Better Teaching Through Testing

Better Teaching Through Testing

A PRACTICAL MANUAL FOR THE PHYSICAL EDUCATION TEACHER

M. GLADYS SCOTT, PH.D.
Associate Professor of Physical Education, State University of Iowa

ESTHER FRENCH, PH.D.
Professor of Physical Education, Head Department of Physical Education for Women, Illinois State Normal University

A. S. BARNES AND COMPANY, NEW YORK

This book has been manufactured in accordance with the regulations of the War Production Board.

PRINTED IN THE UNITED STATES OF AMERICA

Preface

Physical Education has reached the stage where individual diagnosis is recognized as an essential step in the teaching process. The mere recognition of the importance of testing, an integral part of individual diagnosis, does not guarantee the translation of the idea into practice. All too often tests have either not been available or were too long and complicated, too extravagant of class time, or had no proven value. A simplified presentation should be helpful for the teacher who has not specialized in testing procedures.

The problem of testing has been accentuated by war time conditions: by the necessity of showing results, and for obtaining them quickly, by the necessity of mass testing, by requests for measures of qualities other than specific skills, by the need for standards of performance, by attempts to interpret the significance of inadequate or superior abilities. Such necessities have accelerated the rate in experimentation and use of tests, and have promoted an analytical attitude toward testing problems.

It is the purpose of this book to give a non-technical discussion of the testing procedure. It is designed as a practical manual to be used by the teacher in planning various units. The professional student should find it useful as a text, cut down to essentials, and aimed at giving a clear perspective on testing as a part of the teaching procedure. Our thesis is that testing must be considered and used in relation to aims and objectives and that its value is dependent upon the benefit derived by the individual pupil.

All the tests included here have been validated on some group, have been used with other groups, and are known to be reasonably satisfactory. The majority are of recent origin and many have not appeared elsewhere in the literature. No claim is made that they are the final answer. New measures are constantly being constructed, some of which will undoubtedly be improvements. How-

ever, these are practical and have proven value, and the student or teacher can use these suggestions as a basis for evaluating new tests as they appear.

This book aims at giving a background for test construction, selection and use. This background is practically identical for those conducting physical education programs for either boys or girls. A few tests described herein are for girls only. Many others have not been adequately tried with boys' groups, but might be found useful with minor adaptations. Few tests are given specifically for boys. However, the background provided here should be helpful in working with tests from other sources.

Acknowledgments are made to our many students who have served as subjects and often experimented with us on different procedures. Special recognition is also given to our graduate students who have taken the initiative in developing certain tests. References are made to their studies throughout the book. They include the following: Jean Bontz, Ruth Buchanan, Marian Conelly, Bernice Cooper, Dorothy Mohr, Elizabeth Salit, Evelyn Shaufele, Margaret Schmithals, Jeannette Smalley, Wilma Kerr Smith, Evelyn Sturtz, Audrey Underkofler, and Marjorie Wilson. Appreciation is also extended to Marjorie Wilson for her aid in statistical work.

We are also indebted to the publishers of various articles and books for their permission to use some of their material. Recognition is given in the text for such source materials, but acknowledgments are hereby made to the following publishers: A. S. Barnes and Company; F. S. Crofts and Company; Public School Publishing Company; Research Quarterly of the American Association for Health, Physical Education and Recreation; and the Visual Education Department, University of Iowa.

M. G. S.
E. F.

Contents

I.

Use of Tests

Physical education, in common with all branches of education, has certain aims and objectives. A list of objectives for pupil development through physical education usually includes something like the following: health or physical fitness, motor skill, knowledge or information, and social adjustment. If such concepts are definite enough to be set up as teaching goals, if acquisition of learning takes place, then the results must be recognizable and more or less precise means of evaluation must be possible. It is obvious that there is variation in the ease with which the aspects of these different objectives may be measured. It is the purpose of this book to provide practical information on the areas which may be measured most satisfactorily; that means tests for many of the motor skills, for knowledge, and for some phases of fitness not covered in the medical examination.

MOTIVATION

Probably the greatest source of satisfaction in any learning situation is the feeling of accomplishment. Both the teacher and the student share in this sense of achievement. This is one of the reasons that activities such as swimming or stunts gain almost immediate enthusiasm; results are apparent at once. Any activity can be taught so as to secure these same obvious results if practice and self-testing devices are used. For example, does a player use a shot from specified points around the basket with increasing accuracy? Is the softball thrown from base to base with greater accuracy? Is a player's tennis stroke improved so that the number of continuous returns made by the player and an opponent increases regularly? Such systems of more or less objective evidence may be a part of regular practice and serve as a real source of motivation.

The occasional student who lacks that natural interest in self-testing of skills will be motivated by the knowledge that tests will be given and that the teacher and other students will judge him by the skill he demonstrates in these tests.

STATUS OF STUDENTS

There are certain times during the learning period when it is desirable to know the relative status of the students on acquired skills or innate capacities which are pertinent to learning of that activity. At the beginning of a sport season it may be helpful in sectioning classes or selecting the personnel of squads within a class. Most classes can not be made up with individuals of highly homogeneous ability; but teaching and competition are more satisfactory when the personnel of the teams or squads has been determined by objective measures. Such teams may be matched for ability, or poorer students separated for special help according to the plan of the teacher.

Again at the end of a unit it may be helpful to know the status of students because they usually are interested in their own improvement. Likewise, in most schools the teacher is expected to give each student a grade. Grading is a process of rewarding a student according to his learning relative to the learning of the other students and relative to a subjective standard established by the teacher. Since the standards for the extremes of the grading scale must remain more or less subjective, objective measures help considerably in differentiating within this range. They also serve as a check on the justifiability of the teacher's expectations for a particular group of students, or for individuals in terms of capacity.

Further evidence of the level of individual or group performance may be secured by making comparisons with performance of students in other schools. Such comparison can be obtained best by objective tests. This purpose is claimed for interscholastic competition in sports, but such competition usually affects only the very few at the top end of the scale. Of course, there are advantages and disadvantages both in competitive sports and in standardized tests and achievement scales. However, if the *whole* group is to be considered, either by way of determining level of ability or

stimulation to greater effort, then the testing method is preferable.

If the teacher has the same students in similar activities on successive years, economy in testing time may be achieved by careful selection of tests and by saving the scores. Thus the scores obtained in determining ability at the end of the season may be used at the beginning of the corresponding unit the following year. This will be accurate in classifying students the second season providing all have had equal experience in similar activities during the interval. It will be adequate for relative status. Probably such a procedure will be satisfactory as a basis for measuring the amount of improvement during the second unit. However, smaller improvement may be expected from all students. This is due to the fact that students may be "out of practice" on initial tests and, therefore, score slightly less than they did at the end of the preceding season.

DIAGNOSIS

The purpose of diagnostic tests is to single out skills which need special attention during instruction. If a choice is to be made in materials taught, test results may furnish evidence for that choice. Whenever there is opportunity for any individualized instruction and practice the diagnostic test is a prerequisite.

The diagnostic test becomes a practice test if the student repeats it regularly with the definite aim of attempting to raise the score. If individuals with similar needs are put into one squad then the squad may work together. Whether working singly or in a squad each person may be practicing on the skill in which he has demonstrated the greatest weakness.

Most tests for measuring status or achievement of the student or for motivation sample a few of the skills used in the game. The diagnostic battery usually includes a fairly comprehensive list of skills, in order to be of the greatest help to the teacher in planning the course and of greatest help to all students. On the college level the diagnostic test is used very extensively in advising students, determining their requirements and the scope of their electives.

EVIDENCE OF RESULTS

There are times when every teacher is challenged to show evidence of results of teaching. In some cases that challenge is from a principal when his cooperation is asked for physical education. In other cases it may be parents or their representatives, the school board. During the years of educational reorganization and planning following the war this attitude will be more prevalent than ever before. Everyone is asking, "What can physical education do for the fitness and for the education of the students?" They will be asking for information and asking with an open mind, ready to acknowledge merits in a good program.

An opportunity has been handed to the physical education profession by war conditions and augmented by the attitude of those administering the training programs of the armed services for both men and women. All branches of service have devoted an unprecedented amount of time to conditioning, have given more care than ever before to planning of those programs, have made greater efforts to encourage physically active types of recreation, and most branches have attempted to place such programs in the hands of personnel trained in physical education. All branches of military service have made extensive use of tests to determine aptitudes and capacities, measure learning, determine stamina, and in general to promote or reject individuals for specialized service. These factors will contribute to a more favorable attitude toward objective tests and broad physical education programs.

Teachers may be asked by students to show evidence of results. The poor student may say, "I haven't learned anything, I might as well quit trying." The best answer to that is a test sufficiently easy that it will register performance, discriminating enough to indicate small increments of ability, and on a scale which seems to reward slight effort. Then by repeated use of this test the results become apparent. It has been shown that college students at the low end of the motor ability scale show great interest in their ability and derive considerable satisfaction from modest amounts of improvement when testing procedures are used in connection with instructions.[12] For example, they may be learning

[12] See reference in bibliography at end of this chapter.

the principles of effective jumping while using jumping in games such as basketball and volley ball. The tests which are used from time to time may be the vertical jump, or jump and reach, or a broad jump. These are tests which may be objective, are administered easily, and afford an opportunity for the student to apply the principles which have been learned.

Likewise, better students may unconsciously express a need for substantial evidence of their ability when they say, "Oh, I can do that, I don't need to practice." For such students test scores are usually surprising. They fall below the standards which they themselves expect, or below scores achieved by other members of their group whom they would like to surpass, or below the standards shown in achievement scales. Many students are automatically stimulated by such comparisons.

In some cases the teacher himself may debate the merits of different methods or parts of the program, or doubt the adequacy of results. Objective evidence through tests is the most satisfactory answer to this wholesome criticism by the teacher.

INFORMATION AND UNDERSTANDING AS A PART OF PHYSICAL EDUCATION

Most laymen and some teachers consider physical education to be a matter of physical activity, perhaps developing motor skills and contributing certain physiological results. Factual content, understanding and appreciation, attitudes, and general intellectual growth are not always recognized as an important part of the process. When these are accepted as desirable outcomes and planned for in teaching, they must be tested. Since these are objectives common to all forms of education, other educators have pointed the way to testing. This is done principally by written, objective examinations. The success depends upon the skill with which the teacher constructs the examination. The value for the student may be the same as for the skill tests.

TESTING AS A PHASE OF TEACHING

Tests may be chosen for a given purpose but what is to insure their effectiveness? The tests themselves, though inherently good and administered according to standard instructions, will not guarantee effectiveness. The guarantee lies principally in observing the same rules that constitute good teaching. The good teacher presents his activities with enthusiasm, confidence, and faith in what he is doing. The students will respond with enthusiasm and interest. The whole atmosphere and all results are in contrast to that created by a teacher who teaches an activity because it is easier than some other one he might choose, or because he knows that many other teachers teach it and assumes, therefore, that it must be acceptable. The same situation exists in lessons which include testing. The tests which are used will seem important and interesting to the students if the teacher understands and appreciates their value and demonstrates confidence in them.

Every test which is given should serve the student in some way just as every activity included in the program should be planned with the student's welfare uppermost in mind. If the student is to receive greatest benefit and really understand the test, his score must always be made known to him. Not only should the student be told results, but he should be informed as soon as possible. If too much time elapses he loses interest in that particular test and may acquire an attitude of indifference for future tests. Also, if the student is to plan future work on the basis of test results no time should be lost. Information concerning results also implies enough interpretation of those results that he can really appreciate his own status, plan intelligently, and recognize accomplishment when it does occur.

Following up individual needs calls for carefully planned teaching. Tests serve as one of the best sources of outlining needs and should be recognized thus by the student. This is an effective means of motivating an interest in learning activities. Effort is greater when the test is accepted as a personal aid rather than something for the teacher's benefit. Testing becomes important for the individual; he sees a means of learning faster, and therefore, the whole physical education program appears more significant.

Test results should be studied carefully and recognition given for good performance or improvement. Likewise, encouragement can be given the student who is known to have worked hard but who registered only moderate improvement on the test. If attention is concentrated only on the cases which show great gains many students fail to derive the benefits they should from the use of the test. It must be kept in mind that improvement at the upper limit of the scale is much slower than at a lower level. The superior performer usually secures his share of instruction in preparation for competitive activities but seldom does he secure adequate attention and motivation in the regular physical education class.

Wherever any plan of individual selection of activity is in operation some form of counseling or guidance is essential. Such guidance can depend to a considerable extent on test scores. The student can be guided into the activity where instruction will be at his level, and where weaknesses can be overcome. On successive occasions progress can be charted and encouragement or stimulation given. If guidance extends to related recreational choices he can also be led into activities where he is most apt to be successful and happy. By such planning and the relating of testing to practice and playing the student sees the relationship between testing, learning, and his ultimate goal of enjoying activities, or playing on a team. Colleges have used this system extensively not only with major students but in planning for all college students. Similar plans can be used with high school students even though no choice can be given the student of section or activity. Choices may be offered within the hour, squads may work on different skills all related to the same activity, different self-testing or practice tests may be used for different squads, and always the student may be given opportunity to measure achievement.

Tests which put an emphasis on individual improvement almost always appeal to the older children in the elementary school and junior high school. For that age tests serve as a form of competition and an opportunity for individual recognition. Therefore, the tests should be varied enough that the same students will not always be in the upper brackets. In order to make use of this competitive urge to the fullest, recognition should be made to the whole group of the individuals with especially high scores, or those

making greatest improvement, or the squads with the best score. The use of achievement scores for comparison with other schools or similar groups adds zest and competitive spirit to the testing. With older groups this is not as potent a factor but it does continue to operate in some cases.

The teacher must consider the proportion of time which can be devoted to testing just as he must consider the proportion to be devoted to each unit; the proportion between drill and actual participation in the activity; the proportion between discussion, verbal description or moving pictures and actual practice. In every case there are advantages to be derived, but all objectives must be kept in mind and the relative merits of different approaches or proportions weighed. Likewise, testing can provide some real values, but while a greater time allotment may permit a more extensive testing program it probably will not stay in the same proportion as in a smaller time budget. There may even be situations in which the time allotment is so brief that testing may be reduced to a very low point. The most important rule to follow is that every test should serve the student. If no time remains after the testing is done, no follow-up can be made on it. If too little testing is done, inadequate information is available for teaching.

Testing is a device for teaching and learning. The successful teacher is the one who knows the best devices and aids to learning and who uses them skillfully. Under the direction of such a teacher tests fit in so smoothly and naturally that they are accepted by the students, are used by the student and teacher, are understood by the other teachers, are appreciated by the parents, and actually become inseparable from the teaching-learning process.

SUMMARY

The above discussion has classified tests according to general use in relation to teaching. Usually tests have been classified according to specific purpose or ability concerned. Sometimes this classification has been rather remote from the teacher's problem of improving instruction, and often confusing because it seemed to call for too wide a variety of tests. It is possible for a single test to serve more than one purpose. The main values to

be derived from tests, as discussed above, may be achieved through different types of tests. The types named here are from the popular classification.

MOTIVATION. If tests are to be used to increase interest and effort the teacher may select *achievement* tests of general or specific ability, *written* tests, *diagnostic* tests, or *practice* tests.

STATUS OF STUDENTS. If the teacher needs to know the relative status of students or the ability of a class, he may select from the so-called *achievement* tests and use an *achievement scale,* the *classification* devices, the *motor ability, motor capacity, motor educability,* or *written* tests. Also if he is concerned with some special ability he may use the *cardiac functional, strength, endurance, physical capacity,* or *orthopedic* test as the case may require. The information on status enables the teacher to organize class work, give individual help and guidance, determine pupil progress, and assign grades.

DIAGNOSIS. For diagnostic purposes the teacher may use any of the types listed under status; for status he is most apt to be concerned with a single type. In diagnosis, scope and comparison are essential. Therefore, he needs several forms and a scale on identical units where direct comparison is possible.

EVIDENCE OF RESULTS. Again the selection of tests may be from almost any type. The determining factor will be the objectives which the teacher emphasizes and the information which is desired.

For successful use of tests the teacher must start with a clear understanding of his objectives, evaluate and select tests carefully, administer accurately, and follow up the scores for greatest benefit to the student.

BIBLIOGRAPHY

1. Bovard, John F., and Cozens, Frederick W.: *Tests and Measurements in Physical Education.* W. B. Saunders Company, Philadelphia, 1938, Chapter I.
2. Carpenter, Aileen: "The Future of Tests and Measurements in the Elementary Schools," *Journal of Health and Physical Education,* 15, November 1944, p. 479

3. Cassidy, Rosalind, and Kozman, Hilda C.: *Physical Fitness for Girls.* A. S. Barnes & Company, New York, 1943, Chapter 10
4. Howland, Amy: "Contributions of the Achievement Tests to the Athletic Program," *Journal of Health and Physical Education,* 10, April 1939, p. 214
5. Kistler, Joy W.: "A Study of the Results of Eight Weeks of Participation in a University Physical Fitness Program for Men," *Research Quarterly,* 15, March 1944, p. 23
6. Krakower, Hyman: "Testing—Overdone or Underdone?" *Journal of Health and Physical Education,* 13, May 1942, p. 285
7. Larson, Leonard A., and Cox, Walter A.: "Tests and Measurement in Health and Physical Education," *Research Quarterly,* 12, Supplement, May 1941, p. 483
8. Lee, Mabel: *The Conduct of Physical Education.* A. S. Barnes & Company, New York, 1937, Chapters 9 and 15
9. McCloy, C. H.: "Intelligent Use of Tests," *Journal of Health and Physical Education,* 8, May 1937, p. 297
10. ——: *Tests and Measurements in Health and Physical Education,* Second Edition. F. S. Crofts and Company, New York, 1942, Chapter 24
11. Reichart, Natalie: "School Archery Standards," *Journal of Health and Physical Education,* 14, February 1943, p. 81
12. Salit, Elizabeth Powell: "The Development of Fundamental Sport Skills in Freshman College Women of Low Motor Ability," *Research Quarterly,* 15, December, 1944, p. 330.
13. Segel, David: "An Appraisal of the Influences of World War II on Testing Practices," *Education for Victory,* 2, April 20, 1944, p. 3
14. Scott, M. Gladys: "The Use of Skill Tests," *Journal of Health and Physical Education,* 9, June 1938, p. 364
15. Spindler, Evelyn: "Do You Grade or Guess," *Journal of Health and Physical Education,* 2, October 1931, p. 26
16. Wilbur, Ernest A.: "A Comparative Study of Physical Fitness Indices as Measured by Two Programs of Physical Education: The Sports Method and the Apparatus Method," *Research Quarterly,* 14, October 1943, p. 326

2.

Characteristics of a Good Test

Fortunately in some areas there is a possibility of choice between tests. The question which then presents itself is, "What are the inherent qualities of a test which make it useful or really good?"

The criteria most frequently considered pertinent on this point are more or less uniform. Usually evidence on reliability and validity is presented. Objectivity is commonly assumed in construction of the test and some consideration is sometimes given to economy of time in administration. However, most of this evidence is statistical in nature and requires knowledge both of statistics and of tests in order that the reader be fully cognizant of the implications.

There are criteria other than reliability and validity by which the teacher may study prospective tests. These supplement and explain the first type of criteria and make application of principles well known to teaching. They will be discussed first.

TESTS SHOULD MEASURE IMPORTANT ABILITIES

The ability required in any test should be the same as that required in the more general performance which is being tested. This point seems very obvious and yet it is sometimes ignored in the concentration on statistical treatment. It is more clearly explained by some examples from sports. A sports test is usually given to measure general ability in that sport or specific ability in one technique necessary in that sport. The significance of a specific ability for success in playing that game is dependent upon its relative importance with all techniques required in the game.

For example, the mistake is sometimes made of using a test

known as the throw-in as a measure of soccer ability. This is largely because "statistical evidence" has been presented to indicate that it does measure soccer ability. Truly enough, the throw-in will correlate rather highly with general ability in soccer. It would be equally true with almost all sports. The reason is that throwing is a general basic skill, and throwing tests are an important part of any general ability battery. If one studies the game of soccer, it is evident that a throw of this type is used only occasionally by the goalkeeper and by the right and left halfbacks or fullbacks when the ball goes out of bounds. Other than that the players are not allowed to use their hands on the ball. It should be apparent then that for purposes for which sport tests are used that logical as well as statistical considerations should govern the selection. A very similar limitation is found in hockey on the use of the roll-in, yet the roll-in is frequently found in hockey test batteries.

Sometimes an almost identical situation is found in basketball testing. Some test batteries use the throw for distance as a measure of basketball ability. Again it is an excellent measure of general ability. Since players with high general ability make better basketball players than those with low ability, it is obvious that high correlations can be obtained. However, a good basketball team seldom makes use of long passes so it hardly seems desirable to tell the players that their ability as basketball players will be rated by a distance throw.

In selecting tests from various sources the teacher should follow the same procedure that must be used when creating test batteries. That procedure is to outline the skills used and then devise tests each of which uses one or more of those skills.

Two of the sports mentioned above will be used to illustrate this step. Soccer playing requires primarily (1) ability to play the ball and (2) ability of the player to move quickly with good weight control. Ability to play the ball can be subdivided into passing, dribbling, goal shooting, blocking, or trapping. In almost every case this means using the feet for that purpose. Ability to move quickly can be subdivided into running: running with the ball, *i.e., dribbling;* change of direction; and footwork for kicking, feinting and dodging. As one further studies these abilities it is apparent that passes are rather short, with the exception of place

kicks which are sometimes long; that passing is usually done immediately after receiving the ball, and while the player is moving; that goal kicking is almost invariably on a moving ball at some distance from the goal, with more or less interference between the kicker and the goal. Likewise, weight control is inseparable from ball handling, running must be fast but for short distances with quick changes of direction. When this type of analysis is made, the teacher would have little reason to select the throw-in, or the place kick at goal, or the straight dash or straight dribble.

A similar analysis could be made of basketball. That would make apparent the fact that passes are usually short, and made quickly after receiving the ball; that shooting is usually from rather close range, and almost always must be done while moving; that running is rapid but changes of direction are frequent and abrupt. A careful analysis of the game would doubtless be followed by rejection of such tests as the throw for distance, or standing and throwing at a stationary target on the wall, or free throw shooting.

TESTS SHOULD BE LIKE GAME SITUATIONS

The tests should be as nearly like the game situation as possible. A service stroke is always taken from a stationary position and usually with no great pressure of time. Therefore, a test of serving ability may be set up very successfully as successive trials from a given spot at a given target. However, the later strokes or returns in the game are made on an approaching ball, which may require considerable footwork and must be timed with the approach of the ball. It is seldom that such a play can be slow and deliberate. Play is continuous and the test should be continuous.

One fallacy too frequently put into practice, and also into print, is that when skills are combined in a test, it makes the test game-like. That assumption results in such basketball tests as a pivot and shot, or a bounce and shot, or soccer tests such as the dribble and goal kick. Each is scored by the percentage of successful goals. This ignores the element of speed which is almost always present, and the possibility of poor form or actual infringement of rules which may take place. The latter is especially frequent in the basketball bounce and pass test. Even if violations are considered it means

an extra trained helper for each player taking the test and the decision of that helper is subjective. If test situations can be set up where the player makes his own footwork fit the situation and continues the play the results are usually more satisfactory.

The time element is much more important than is usually realized unless a systematic study of the problem has been made. Let us consider an example from softball. Successive throws from a standing position at a stationary target proved very poor. A ball sent from a catapult, caught, and thrown immediately at a target proved somewhat better. However, that offered two difficulties. First, the use of a catapult is impractical. Second, it was still possible for the player to hesitate long enough after the catch that the situation was almost the same as in the stationary throw. The next step in the development of this test was to change the sequence and add a time control. The player then threw the ball at the wall, fielded it as though playing a baseman's position, and threw to a target which represented another baseman. Time was counted from the start of the ball on the first throw to the hit of the second throw. Accuracy of the second throw was recorded. The actual time required for those throws proved relatively unimportant except in the extremes, but it brought the throwing into the same timing as experienced in the game. Therefore, performance on the test corresponds much more closely with actual performance in the game.

Timing in the above test is one of the few examples where records are taken and no very definite follow-up made of results. However, it has served its purpose of making the test psychologically game-like. The difference in the two forms of the test is comparable to the difference between pegging the ball around the bases for practice and trying to beat the runner to base for an out.

TESTS SHOULD INVOLVE ONLY ONE PERFORMER

The above discussion has doubtless suggested the use of two or more players in a test situation since the player in the game must always consider the person from whom he receives the ball, to whom he passes, and very frequently his opponent's actions. Such an arrangement would satisfy the standard of game

similarity. However, no one standard can be considered in isolation from the rest. The cooperation or competition for players A to Z must be identical when being tested. It is perfectly obvious that they must all have good balls, the same size target and other equipment. It is equally obvious that one player should not be tested with a partner who fumbles and passes wildly while another has an excellent player.

The test should involve only the one person being tested. This accounts for the very frequent use of repeated throws or volleys against a wall. The objection is that this type of test presents an artificial situation. This is true. However, it represents a compromise between the two criteria of game similarity and a single performer. A repeated volley test does aid in a way in producing game similarity in that the player both receives and passes the ball, and play is continuous. Moreover, the player alone is responsible for the results.

TESTS SHOULD ENCOURAGE GOOD FORM

Another problem of real importance in some sports is that of measurement relative to form. For example, a tennis player may be able to place a ball in a specified area on the court but the flight of the ball may be very slow, arched and followed by a high bounce; the return of such a ball would be extremely simple. Another may send the ball in the same area but it travels with speed, in a flat path with little bounce. Some tests do little to distinguish between these two players.

The best solution to this difficulty would appear to be either the introduction of the time element or a subjective rating of form to supplement the test score. The first approach is used in a test such as the Dyer backboard test (see p. 100). However, this will ignores form and as usually given does little to eliminate that objectionable characteristic. That can be partially overcome by moving the restraining line back to 25 feet (perhaps with the privilege of stepping across for a single stroke to be followed by another from behind the line). That prevents a player from standing near the wall, volleying the ball, and receiving a high score because the ball travels only a short distance. When everyone plays

from a more nearly uniform position, the player with power and control has opportunity to score above inferior players.

TESTS SHOULD PROVIDE ACCURATE SCORING

The objectivity of a test depends upon the certainty with which the trial can be scored a success or failure, or for a given value on a target, stopwatch or measuring tape. A basket with a net leaves little doubt as to whether the ball went through or dropped outside. A badminton bird which is to be scored as going above or below a rope is more difficult to see. A ball thrown at high speed at a target is difficult to judge as inside or outside a line unless special equipment is used for that purpose.

A simple form of target to construct for accuracy throwing is made of wood. The center of the target, and alternate circles out from the center, are made of tin pieces attached to the wooden background. The sound of hits in two adjacent areas makes it possible to definitely judge the accuracy of the hit.

Distance covered by a basketball bounce and the legality of a play in terms of traveling are difficult to judge. When a race is judged by the zone in which the runner finished, errors in judgment also creep in. The faster the runner, the greater the chance of error. If given indoors on a course adjacent to a wall, the numbers can be placed on the wall instead of the floor.

TESTS SHOULD PROVIDE A SUFFICIENT NUMBER OF TRIALS

Trials should be sufficient to eliminate chance deviations from a truly representative score. One trial may be sufficient under optimum conditions. For example, a race is usually set up with uniform conditions for all. Assuming a good timer and proper motivation, an accurate measure of maximum effort can be obtained from a single run. Ability on short runs will not vary much from day to day, ability on longer endurance runs might vary more, but they are usually used for a different purpose.

The number of trials necessary for a given test can be determined only by experimentation. Some general rules can be stated, however. Most tests of maximum effort can be measured best by

one to three trials. This would include items such as a dash, throw for distance, strength events, and speed events where control is relatively unimportant. When a high degree of accuracy is necessary the number of trials required goes up. It may run then from five to thirty, usually between ten and twenty. Frequently the number of trials necessary for a group of advanced players will be fewer than for less experienced players on the same test.

If the test is being used for a practice device or for motivation, the trials may be reduced more than if the purpose is classification or grading. For example, it may be desirable to have some players practice on free throws. As a practice device five to ten may be the maximum number of throws that is practical; and in this form a test may be valuable as a motivator as well as providing practice. However, it is known that this is an insufficient number for reliable estimates of a player's ability.

TESTS SHOULD BE INTERESTING AND MEANINGFUL

The test should appeal to the students if best efforts are to be obtained. This is partially a problem of administration. The means of motivation are varied: using individual score cards with continuous records; posting scores and names of the best performers, or of those who make greatest improvement; promotion to better teams or squads when sufficient improvement is made; or comparison of scores with accepted standards. However, there are certain inherent qualities which attract or repel student interest. Game similarity is the first step toward a favorable attitude; the test then has meaning.

When a student knows how well he is doing on the test, or at least what the score is when the test is finished, he is more interested. Probably, one reason for the popularity of the basket shooting tests is the fact that the student can see immediately how each attempt is scored and each successive shot presents a new chance under identical conditions with the first.

Different types of scoring carry different meaning to the students. For example, some tests may be scored either as the number of hits in a given time interval or as the time required to make a

specified number of hits. A throw or hit is perfectly understandable to the student. However, when results are expressed in time intervals, number of seconds, it is a rather vague concept. It is true that the student can make comparisons with others even if he does not fully understand the score, but the standard of an optimum or ideal performance is lacking.

Test should not be so time consuming that the student associates them with slow moving class periods when he is deprived of the opportunity to play the game. The tests should fit into the time that is ordinarily devoted to practice on techniques.

Most sports tests do not produce any undesirable after-effect. However, some types of tests which are done for endurance may result in severe muscular soreness. Few girls delight in that soreness or even take a stoical attitude toward it. The result is a dislike for the test which produced it, or perhaps for all tests, because they do not understand the causes of the soreness or the types of tests which produce soreness. Boys are less apt to be affected by this factor.

DIFFICULTY OF THE TEST SHOULD BE ADJUSTED TO THE ABILITY OF THE GROUP

A test may be easy or difficult according to the way in which it is done. Tests may need to be modified to meet the ability of the group being tested. It is not recommended that tests be re-made freely since one can no longer assume constancy of desirable characteristics in the altered form. However, strength tests may be made easier by changing the kind of support, runs may be shortened to the point where students will keep trying throughout, and throwing distance may be shortened so that all can throw as far as the target.

The rule to be applied here is that scores should show a reasonable distribution. This means there should be no massing of scores at any one point. If the test is too easy, most of the scores will be at the maximum. If the test is too difficult most of the scores will be zero or near it. In such cases the form of the test may be changed. However, it is better either to postpone the test until the class has developed the necessary ability, or to select another test.

TESTS SHOULD BE JUDGED BY CRITERIA WHICH CAN BE APPLIED STATISTICALLY

How do all of the above considerations relate to the usual textbook list of criteria, namely: validity, reliability, objectivity, economy? Reliability, or consistency of measurement of the same degree of ability, is possible only if the performer is interested enough to give maximum effort, if the trials are sufficient in number to eliminate chance factors, if his own skill only is measured, if equipment and test conditions are uniform, and if measuring units are objective.

Validity is defined as the degree to which a test measures what it is supposed to measure. As stated before, *logic* as well as *statistics* should be applied in determining that characteristic.

Economy results primarily from careful selection of a few highly valid tests which are administered efficiently (see Chapter 3).

Norms or standards of achievement undoubtedly add to the usefulness of the test. But that is not an insurmountable obstacle. If achievement scales are not available, a substitute scale may be constructed on a given class or school group. This is most easily done by the T-scale (see p. 222). In some cases this may be preferable to using published scales since the standards are then in terms of that particular group.

If one is to make use of the work of others on test evaluation certain statistical concepts are essential. Frequency distributions, measures of central tendency and of dispersion, product moment correlations and regression equations are formidable sounding terms. A speaking acquaintance with the terms as a handle for the concepts they represent is adequate for evaluation of testing reports.

Basic to most of these concepts is that of the range of ability or distribution of scores. In any large group, unselected for the ability in question, there will be considerable variation in the level of that ability. If the measuring device is sound the range of ability will be shown. However, there will not be an equal number of individuals at all points along the range. In fact there will be only a very few cases at each extreme. Points toward the center of the range show an increasing number of cases. When put into graphic

form it is a symmetrical "bell-shaped" curve. In other words individuals tend to be alike in ability, but there are always variations toward superior and inferior ability. The measuring device must discriminate between ability at all levels. It is in the middle of the range, or in the area of similarity, that discrimination is apt to be inadequate. The achievement scale is simply an elaboration of the distribution of ability shown by a given group on a given test.

This last statement then suggests that distributions vary for different groups. Since this is true, one is justified in expecting every report of test construction to describe the population on which the report is based. That is, it should state the age or grade level; the sex; number of cases; and special characteristics, if any, which would affect the results.

The reliability and validity of a test are expressed as correlation coefficients. Such coefficients are merely a numerical expression of a degree of relationship. The reliability coefficient expresses the relationship between two consecutive administrations of the test, therefore, indicates its consistency of measurement. It is not always practical to give two duplicate administrations when studying tests. As a substitute on tests where several trials are given, the alternate trials may be correlated (the so-called odd-even method of determining reliability). This certainly measures the consistency within the series of trials. It must be remembered that such a coefficient will usually be lower than if two administrations had been given, because the correlation is actually made on half of the series of trials. Correction is possible for that difference and is sometimes made. The method of correction most commonly used is by the Spearman-Brown Prophecy formula. The reader should note in every case whether such correction has been made in order to know how to interpret the coefficient.

The validity coefficient expresses the degree of relationship between a criterion and the test. The higher the relationship, the more truly does it measure the ability in question. It is obvious since the criterion is the yardstick against which the test is compared, that the criterion must be a good one. The reader has the right, therefore, to expect a statement of the criterion used. If ratings are used they must be acceptable (see p. 154).

Since reliability and validity are expressed as correlation coefficients, it is necessary to interpret these quantitative values. There must be a clear distinction between .90 and .09; between +.90 and —.90; also one must understand how much better .90 is than an .80, and why coefficients are always decimal values and never whole numbers.

These points can be understood best by a consideration of the principle and procedure by which the coefficient is computed. The correlation is a mathematical expression of relationship between two factors or abilities as measured on a given group of persons. If it is a reliability correlation, the ability as measured on the first day is compared with a similar performance by the same group on a second day. If it is a validity coefficient, one factor is the criterion by which the measure is being checked, and the other is the experimental measure itself. The correlation is essentially a process of plotting the two factors under consideration into a single graph. This graph is called a *scattergram.* Let us consider Figure 1. The vertical scale OC represents the distribution of cases on the criterion. If fifty cases had been placed in rank order on that scale

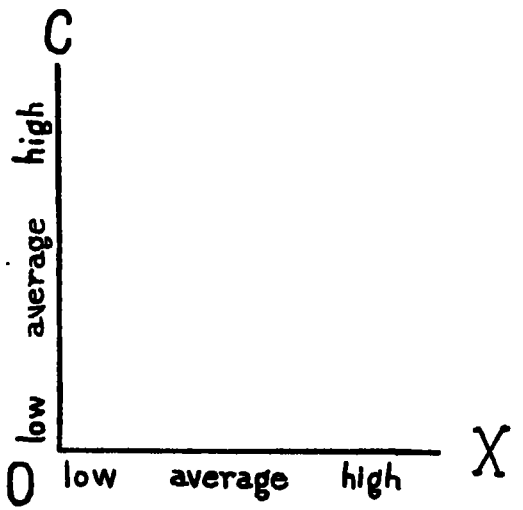


Figure 1. Diagram of Correlation Graph or Scattergram

then #1 is at the top or high end of the scale and #50 at the low or zero end of the scale. All other cases are in successive order between these two. Likewise, each case has been placed in rank

order in the experimental measure to be plotted according to XO from high down to low.

The effects of similarity and discrepancy in these two rank orders is demonstrated in Figure 2. In part *a* it is assumed that the two orders are in perfect agreement. The same person would be in the top position each time and would be represented by #1 on the scattergram. All other cases would be similarly ranked and their placement on the graph would be in a straight line from O diagonally upward. The computation is such that if this occurred the coefficient would be 1.00. Such an instance probably would never occur because of chance variations, differences in effort, errors in measurement and other similar shortcomings of measuring techniques. Occasional slight variations in rank may occur and the plotting on the scattergram would then appear more as in Figure 2, part *b*. Such a scattergram shows the same general direction in the design of tallies across the graph but they do not hold to a narrow straight line. A scattergram such as *b* would give a coefficient of approximately .90, or something a little less perfect than *a*.

On the other hand if there is no similarity between the rank orders a scattergram such as *c* would occur. In that case an individual who is low on one axis may be anywhere from low to high on the other, and the individual who is high on the first may be of any rank in the second factor. Such a heterogeneity of tallies would produce a coefficient at or near zero.

There might be instances where the rank order of one scale is completely reversed in the second. Such a scattergram is shown in part *e*. The tallies all fall in a straight line as they do in *a* except that they run from C to X instead of from O up through the graph. The scattergram in *e* would also give a coefficient of 1.00 but it is distinguished from the 1.00 in *a* by a negative sign. Part *a* shows a direct and exact correspondence of the two scales; its coefficient is called positive (written + 1.00) and it expresses relationship which is direct. Part *e* shows a perfect reversal of the two rank orders; its coefficient is called negative (written — 1.00) and it expresses an inverse relationship between the two measures.

It should be apparent that perfect inverse coefficients are no

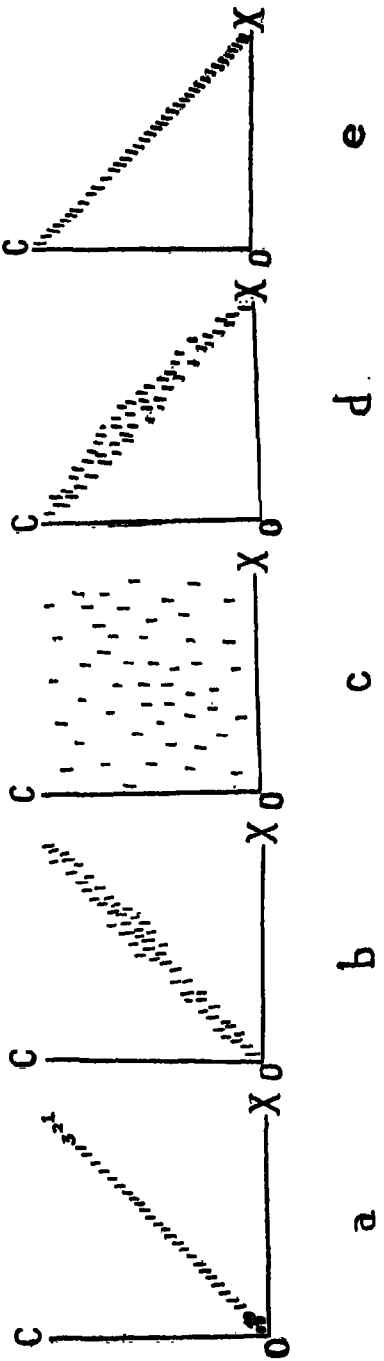


Figure 2. Scattergrams Yielding Different Coefficients

a— + 1.00 coefficient
b— + .90 coefficient
c— zero coefficient
d— — .90 coefficient
e— — 1.00 coefficient

more likely to occur than perfect positive ones. Therefore, part *d* represents the more usual form of a high, inverse relationship.

No exact arbitrary points can be set along the numerical range of coefficients as a point of significance or as a high or low degree of relationship. Generalizations can be made, however. Higher values are necessary for reliability coefficients than for validity coefficients. This is true because the validity coefficient cannot be higher than the reliability of the measure, in fact it is always somewhat lower than its reliability. Reliability coefficients of highly complex skills used in physical education are usually lower than those to be found on tests of mental capacity or achievement. Tests which call for extreme all-out effort, such as endurance tests, frequently will be less reliable than those which use a submaximal effort. Tests on girls are usually less reliable than similar tests on boys, apparently since it is more difficult to motivate the girls to their best efforts. The performance of inexperienced players is usually less reliable than that of highly skilled ones. The reliability of a measure increases with the number of trials, though it is far short of being in direct proportion to the increase in number. A coefficient computed on many cases can be relied upon as being more stable than a similar one computed on very few cases. A minimum of fifty cases is desirable, preferably a hundred or more.

With the above points in mind let us then generalize on qualitative interpretations of numerical coefficients. Anything above .85 is considered very good, but above .95 almost impossible. From .75 to .85 is considered adequate for many purposes. As reliability coefficients drop below the .75 value they indicate an inconsistent and poor measuring tool. A validity coefficient below .60 to .65 indicates poor predictive value.

It must not be assumed that low coefficients are meaningless. Rather they indicate a lack of relationship and that fact should be interpreted accordingly. A low reliability coefficient, indicating inconsistency of measurement may suggest additional trials, or improvement in details and conditions under which the test is conducted. A low validity coefficient would indicate that the test is worthless for predicting the ability which it was assumed to measure. However, if it seems to have some merit as determined subjectively and is shown to be highly reliable it might make a

very good practice test and serve the purpose of securing interest and effort from the pupils. (Free throw shooting in basketball is an example.) Achievement records or grades, should not be based on such a test.

In another case several tests might have been studied for their relationship to general ability in a sport. Each of them may show fairly high validity coefficients, i.e., high relationship to this general ability. Part of the tests may correlate highly with some of the others, and some may show very low correlations with others. If a combination of tests is to be used, then those which have high intercorrelations would be discarded because they measure the same or similar qualities. Those with low intercorrelations would be selected because they measure different aspects of the general ability.

The above principle explains the basis for construction of test batteries, or series of tests, to measure or predict a general ability. The validity of the total battery is computed by a multiple correlation instead of by the procedure just outlined. In order to get a high validity from such a combination it is essential that each test be good, i.e., reliable and reasonably valid. It is also necessary that the intercorrelations be low. Let us illustrate these statements with a rather obvious example. Suppose you have a box; you know nothing about its contents and it cannot be opened. One means of describing it to someone else would be by its shape or objectively by its external measurements. Suppose you use a good cloth tape to determine its length, breadth, and thickness. That would certainly be descriptive but it would add nothing to your further description of the box and its contents to proceed to use a flexible steel tape and measure it again. You would simply have two measures of the same thing, i.e., two measures with high intercorrelations. It would be much more helpful to determine its weight as a second measure. While size and weight may tend to be related, there doubtless would be a lower intercorrelation between weight and size than between size as measured by the cloth tape and size as measured by the steel tape. Knowing weight and size we now have a better clue to the nature and amount of the contents than was known before. And so additional qualities may be measured if the means exist for that measure-

ment, and additional measurement will be valuable if it does not duplicate something already done.

Multiple correlation coefficients are expressed in the same form as those of the simple correlation, but are always positive. Somewhat higher values of multiple coefficients are expected than in the case of a single test since more measurements are being considered.

The multiple correlation tells which tests to combine in a battery to measure general ability but it does not tell how to combine them. Frequently the units in which one test is scored are so different from those of another that one would completely dominate the final score if they are simply added together. For example, you might wish to combine scores from a softball throw for distance where the average score was 60 feet and a target throw where the average score was 15. The score on the throw is always so large that variations in accuracy on the second test would fail to show up in the total score. It might also happen that accuracy was more important in the estimate of ability than strength for a distance throw. The scores are almost always weighted, therefore. This is done by the regression equation which specifies the proportion of the raw scores to be used.

The other alternative to the use of the equation when combining tests is to take the student's T-score for each test and add these together. (See p. 222 for explanation of T-scores.) Both procedures take into consideration the range of all scores and the ability of the student with respect to the rest of the group. The sum of the T-scores is computed more quickly and is very satisfactory.

TESTS SHOULD PROVIDE A MEANS OF INTERPRETING PERFORMANCE

The average and mean are synonymous terms. The median score is that point in the scale that is excelled by fifty per cent of the cases and is better than fifty per cent. The mean and median, as measures of central tendency, are not particularly useful to the teacher who is concerned with individual performance. The measures of distribution (or difference from the mean or median) are much more valuable. These measures are the quartile,

percentile, and the standard deviation. The quartiles locate the scores between which twenty-five per cent of the cases fall, in other words each half created by establishment of the median is again halved. See Figure 3 where each quartile contains twenty-five per cent of the cases and point 2 is the median. It will be observed that the score range for A-1, 1-2, 2-3, 3-Z may not be the same. A student

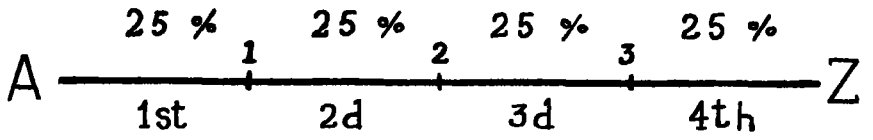


Figure 3. Quartiles

A-Z—range of scores
A-1—lowest or first quartile
1-2—second quartile
2-3—third quartile
3-Z—upper or fourth quartile

understands very readily when told in which fourth of the class his score falls.

Percentiles follow the same plan with subdivisions into smaller than twenty-five per cent. A student also understands readily what is meant if told that his score is on the 65th percentile, i.e., it is better than sixty-five per cent of the group but surpassed by thirty-five per cent.

The standard deviation is a yardstick commonly applied to any symmetrical bell-shaped distribution, approximating the *normal curve*. The curve is bell-shaped because the frequencies decrease as the amount of deviation from the center increases in each direction. Since the standard deviation is a "yardstick," the units on it are the same length throughout the length of the scale. Each standard deviation is named according to its deviation from the center or the mean. Because of that specific location, each has a specified number of frequencies in it (Figure 4), and is equaled only by its partner on the opposite end of the scale. The length of each standard deviation is uniform for each distribution but its length differs for each distribution since the total range varies and the distribution of cases through the range varies. The size of the standard deviation increases when the total range increases, or when the scores fall with greater frequency near the extremes.

The T-scale is an easy and practical interpretation of the

standard deviation of a distribution, and is the basis for many achievement scales. The mean is always 50. Each standard deviation equals 10 points on the T-scale. Therefore, the limits of the scale are approximately 20 and 80. Likewise, it is apparent that about two thirds of the class will have scores between 40 and 60 and that there will be very few 20's and 70's. A simple explanation to the student suffices. For example, 50 is average on the test; above 50 is better than average, the higher the better; below 50 is less than average, the lower the score the poorer it is.

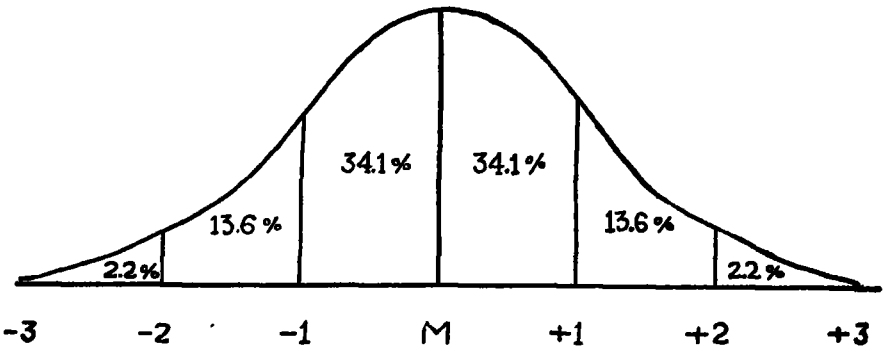


Figure 4. Standard Deviations

M — mean of the distribution
\+ — above the mean
— — below the mean
1 — first standard deviation from mean
2 — second standard deviation from mean
3 — third standard deviation from mean

The chief advantage of T-scores is that a direct comparison of performance on different tests is possible. It is impossible to say whether a 50 foot throw is better than 10 points on a target test unless each is interpreted as good or poor. The T-score does this on the basis of standing with respect to the rest of the group. Such a comparison is invaluable in diagnostic testing and in motivating student effort.

The T-scores may also be added together for a composite score as mentioned previously. There is adequate evidence that relative standings are almost unchanged when regression equations are used as compared to the sum of the T-scores. (Correlations by the authors on such comparisons range from .94 to .98.)

Some achievement scales range from 0 to 100. They are constructed on the same plan except they are based on the probable error rather than the standard deviation. The probable error is

.6745 of the standard deviation, therefore, the distribution has a range from — 5 probable error to + 5 probable error. With one probable error equal to 10 points the range is 0 to 100.

Either type of scale may be used according to the wish of the teacher and whichever type is available, if using scales from other groups. There seems to be one advantage to the T-scale, particularly when used for stimulating student effort. The poor students may have tried hard and actually received some score on the test. If the probable error scale is used the reward may be less than 10 and may be very discouraging. If the T-scale is used the reward is in the 20's. Such a score seems to give encouragement. At the other extreme a student who makes in the 90's is apt to feel satisfied and not make much further effort. On the other hand if the T-scale is used, which places him in the 70's, it seems to point up the fact that there is still room for improvement.

BIBLIOGRAPHY

1. Cassidy, Rosalind: *New Directions in Physical Education for the Adolescent Girl in High School and College.* A. S. Barnes & Company, New York, 1938, p. 124-139
2. Davis, Elwood C., and Lawther, John D.: *Successful Teaching in Physical Education.* Prentice-Hall, Inc., New York, 1941, Chapter 22
3. Glassow, Ruth B., and Broer, Marion R.: *Measuring Achievement in Physical Education.* W. B. Saunders Company, Philadelphia, 1938, Chapters 1 and 2
4. National Education Association, American Education Research Association: *Review of Educational Research,* "Educational Tests and Their Uses," 8, December 1938, Chapters 4 and 6
5. Williams, Jesse Feiring, and Brownell, Clifford Lee: *The Administration of Health and Physical Education.* W. B. Saunders Company, Philadelphia, 1940, Chapter 19

3.
Administration of Tests

There are numerous conditions which affect testing, and since many of these vary for the different tests and teaching situations, only those principles fundamental to all testing will be discussed in this chapter. Suggestions for specific tests and for adaptations to local situations will be found in Chapters 4, 5, 6, and 8. A few conditions which vary from one situation to another are the number of persons to be tested, the time allotment, the amount of floor, field, and wall space, the amount of equipment, and the use to be made of test scores.

CLASS ORGANIZATION

The amount of time that it takes to administer a test can be reduced considerably by careful organization. First, we will consider the organization and administration of motor tests, with knowledge tests coming later in the chapter. Some tests can be administered along the edges of the floor or field, without interfering with the progress of the class activity, while for others it may be necessary to use the entire space for testing during a portion of the class period. In the former situation, student leaders and the squad type of class organization are most helpful. The squads or teams used for class instruction can be the unit for testing. The instructor should prepare instructions and plan the demonstration in advance and then present them to the entire class at one time. For some tests, practice trials are necessary. Care should be taken to see that all receive an approximately uniform amount of practice. All this can be done before dividing into groups. The testing may take a portion of several class periods or one may prefer to devote a few consecutive periods to it. No one plan is best for all situations and for all tests.

USE OF SPACE

A careful study should be made of all available space and plans made for its efficient use. If several persons can be tested at once, the amount of time necessary for conducting the tests is reduced. For example, if a test for accuracy in softball throwing is being conducted, instead of having just one target, arrange for as many as space and balls will permit, being careful to allow enough distance between targets for the scorers to stand without danger of being hit.

Another way to speed up the testing is to arrange areas for practice trials at one side if limited equipment is available for the actual administration of the test. An example of this is the badminton serve test described in Chapter 4. Strings can be tied to a post (standard) placed off the courts, one at net height and another twenty inches above. Marks indicating the floor plan can be chalked or painted on the floor. Oilcloth targets, to be placed on the floor, adapt nicely to this procedure as they can be moved at will. This arrangement permits practice space before going on the court for the actual trials. A similar arrangement can be constructed for practicing the badminton clear test.

TRAINING THE ASSISTANTS

In large scale testing, or where student assistants are to be used in conducting and recording, at least one organizational meeting should be held. The purpose of this meeting is to give everyone a complete understanding of the entire process. The instructions should be explicit and all should be impressed with the necessity of having uniformity of procedure. Definite typed or mimeographed instructions should be prepared and given to each assistant, to be reviewed just prior to the actual testing.

All dimensions given by the test author should be carefully followed, if you expect to secure comparable results. If you are only interested in making comparisons within your own group, then alterations in the tests can be made. Time intervals, amount of practice, and the number of trials should be held constant. Sometimes the wall space is interrupted so much by windows, stall bars,

overhanging balconies, doors, and other fixtures that you can only approximate the dimensions. When this is the case, you cannot be sure that the test results will be either as reliable or as valid as found when administered under more ideal conditions.

Scorers often need special training. This is particularly true when they have to make a split-second decision as to where the object landed. Various devices have been tried in an effort to secure greater accuracy in scoring. Brophy[1] made an accuracy target with four concentric circles, alternating tin with wall space to secure sound effects. (See p. 16 for description of a portable target.) Schmithals[2] placed a wooden board across the goal posts to assist in determining the exact second that the ball reached the goal line. A large target has been used successfully for a volley ball serve test. Some of these special devices are cumbersome to move and require extra time and materials for construction. The scorers should experiment during the training period until they find the spot where visibility is best. For tests involving wall targets, standing to one side and about ten feet from the target works well. The object should be seen in flight and the scorer should shift his gaze directly to the target before the contact. Sometimes it is wise to have both a spotter and a recorder, with the recorder repeating the score aloud before recording, as a further check on accuracy.

The training of assistants is particularly important in events where form is to be judged or fouls called. It is wise to have them go through the tests themselves, emphasizing the correct and the wrong points, and then practice judging each other until the person in charge of testing is certain that they are judging with a high degree of uniformity.

If the assistants are to help in motivating the students to their best efforts, then they should be given hints in doing this. The manner in which the tests are presented is extremely important, and will be discussed later. Giving words of encouragement, having a sincere interest in the individual's score, and showing pleasure at extreme effort are but a few of the ways in which assistants can motivate students. Care must be taken not to embarrass the poorly skilled persons or to make the actual score seem the all

[1, 2] See reference in bibliography at end of this chapter.

important thing. On the other hand, unless the students are motivated to their very best efforts, the scores are meaningless, and the time spent in testing might better have been spent otherwise.

If the test directions call for stop watches, assistants should be given instructions concerning their use and how to read them. All stop watches should be checked by a jeweler immediately before the tests are conducted, if there is any discrepancy between them or doubt as to their accuracy.

DEVICES FOR FACILITATING MEASUREMENT

Targets are used in many tests. A satisfactory method to use in placing them on walls or floors is to paint them. A quick drying, washable paint should be used. Show card or poster paint, which can be purchased in a wide variety of colors, works very well. It can be removed with a damp cloth and lasts indefinitely on wall surfaces or on floors until they are scrubbed. Some colors may leave stains if applied to a porous surface and should be pre-tested. In making targets of concentric circles, a string tied to a piece of chalk can be used to outline the pattern. Knots can be tied in the string, corresponding to the various radii. Care should be taken to select string that does not stretch easily. The directions for tests usually indicate the width of lines; a paint brush of the same width is convenient.

The use of different colors aids in scoring. Another aid is the painting on the target of the score value for each space.

Adhesive tape, torn in proper widths, can be used in constructing wall targets involving straight lines, such as the repeated volleys test (see p. 101) or in a tennis serve test. This is a quick but rather expensive procedure.

If you wish to be able to move a target from place to place, the target can be made of some lightweight material, such as oilcloth. Crayola or colored chalk can be used on the rough side of oilcloth or the smooth side can be painted, or outlined with India ink. This type of target can be used on dirt or grass surfaces, or placed on flooring.

Ropes, strings, or poles are used in some tests to insure certain heights or distances. Two bamboo poles, placed across jumping

standards can be used to outline a target. A clothesline rope is used in the badminton clear test, placed across the floor fourteen feet from the net and parallel to it, at a height of eight feet from the floor. The rope can be attached to high standards, such as those used in the pole vault or for tetherball, provided that their bases are heavy enough so that the weight of the rope will not cause them to tip slightly, thus allowing the rope to sag. Screw rings can sometimes be placed in the walls; in the case of brick walls, the screw can be inserted in the mortar between bricks. The strings used in the badminton serve test can be tied to the standards holding the net. If these standards are not high enough, they can easily be extended by taping lightweight sticks, such as yardsticks, to the standards. Sandbags can be placed on the bases of standards if necessary to hold them erect.

In tests involving throws, kicks, or hits for distance, time will be saved if the field is laid out so that the contestants can start from either end. Lines parallel to the starting line can be marked, thus permitting quite a few to be tested at once, providing the field is wide enough and the supply of balls is ample. Markers can be placed along both sidelines, indicating the value of the zones, progressing from the zero or starting line to the highest score. These markers should be placed along the right hand sideline. (Figure 5) The lines across the field can be made with lime, using regular field liners; or with string or linen tape, tied to sharp objects and pulled taut to the ground. If the number of subjects or trials is small, these lines are not essential. Small objects, such as sticks or stones, can be used to mark the landing spot. If each contestant is to have three trials, with only the best effort being measured, a pointed stick (tongue depressors are convenient), bearing the contestant's number or initials, can be inserted at the spot. Measurements can then be made when all trials have been completed.

In some track and field events, such as the discus throw for boys, and the shot put for boys, the regulations call for measuring directly from the starting point to the spot where the object first broke the turf. Here there is no desire to penalize for lack of accuracy in direction. It is recognized that when zones are laid off by marking lines parallel to the starting line that there is a penalty for gross inaccuracies. To avoid this, for the events mentioned

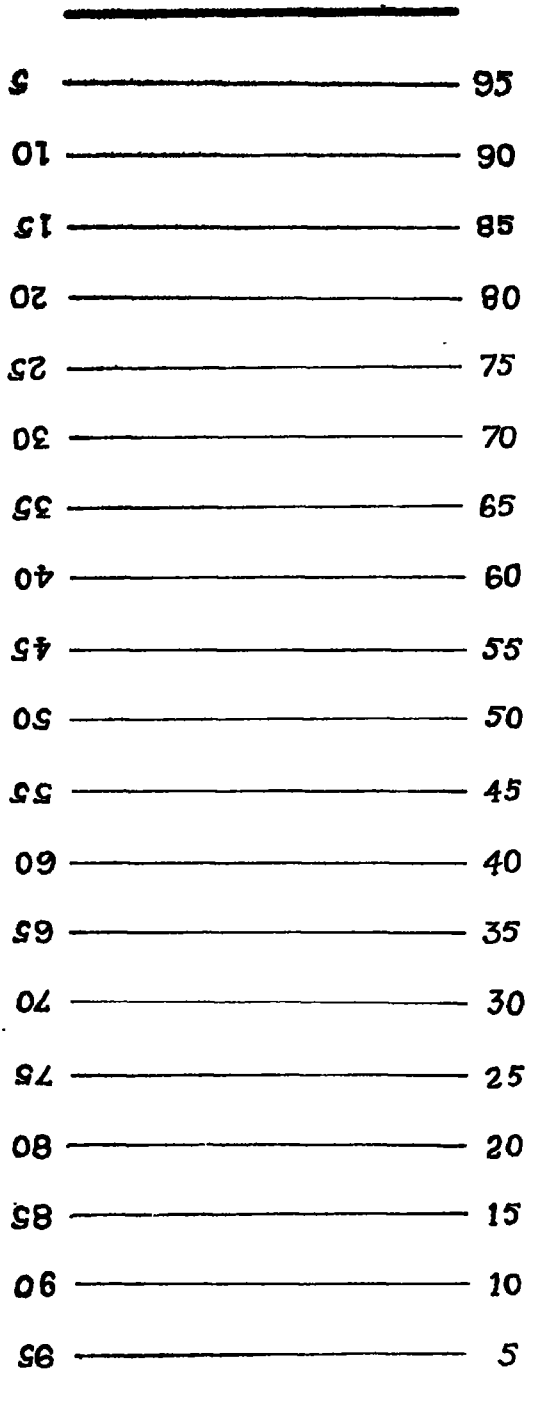


Figure 5. Parallel Field Markings for Distance Events

of time. Thus, one stop watch can be used for the entire group being tested. A spotter is assigned to each runner, and is expected to be opposite the contestant at the end of the time interval. If properly trained in observing running skill, the spotter will be able to make adjustments in position for judging the various speeds. Markers which can be easily read should be placed along both sidelines, so the spotter can score the finish by looking across the lanes. (Figure 7)

Another method for timing several individuals with only one stop watch available, provided the time is to be taken to the nearest full second, is to have the timer, stationed near the finish line, count the number of seconds aloud. The spotters watch the individual runners or swimmers and record the time called immediately before the individual reached the finish line.

When stop watches are not available, a metronome can be used. The sound can be supplemented by drum or by voice. The metronome can also be used in standardizing rate in such tests as endurance tests where a certain exercise is continued for a considerable period of time.

Lines on the mat covers can be used for convenience in scoring such events as the standing board jump. If only a few lines are drawn and measurement is desired to the nearest inch, a ruler may be used to measure from the nearest line to the imprint of the foot.

In conducting an event such as the high jump, where the total number of trials is not predetermined, it is suggested that the group be divided into squads according to ability. If this is unknown, then divide them according to height, sending the taller jumpers to one set of standards and the shorter to another. If only one set of standards or one jumping pit is available, small squads should be used and they should be rotated to various events, in order to eliminate the time wasted while standing to wait for turns.

SCORE CARDS

Score cards should be prepared in advance, with the names of the persons to be tested and the number of trials indicated. Where scores are to be converted into some other form, a

space should be provided for the converted score, adjacent to the total raw score. A box of pencils should be placed in a convenient spot.

Various types of score cards have been used, and each has its merits. If scores are to be used from one quarter or semester to another, then the individual score card, which can be turned over to the new instructor or re-alphabetized in the new grouping, is preferable. A sample of such a card is shown below:

NAME _____
Last *First* *Middle* *Classification*

FITNESS TEST SCORES

Date	Physical Education Class Activity	Sit-up	Bouncing	Chinning	Run

Ht._____ Wt._____ Age_____ Health Rating____ Posture_____

Figure 8. Sample of Individual Score Card

Squad cards are convenient and economical, when the testing is being done with the squad as a unit. Sometimes the card used for taking roll will have sufficient spaces for recording test scores. When the number of trials is five or more, the addition of scores will be facilitated if the card is arranged so that the scores can be recorded in a vertical column rather than a horizontal row.

If performance levels are available for the test, compiled from scores made by groups similar to the group being tested, it is helpful to have them placed on the individual score card. They make

TABLE I

ACHIEVEMENT SCALES FOR COLLEGE WOMEN [a]

Arranged for Construction of Student Profile

T-Score	*Obstacle race* [b]	*Pull* [c]	*Push-ups* [d]	*Sit-ups* [e]	*Bouncing* [f]	[g] *T-Score*
81			41-up	108-up	210-up	81
80			40			80
79	17.8-17.9	86	39	102-107		79
78			38	96-101	200-209	78
77		85	37	84-95	190-199	77
76	18.0-18.1	84	36	72-83	170-189	76
75	18.2-18.3	83	34	70-71	160-169	75
74	18.4-18.5	81	32	68-69	157-159	74
73	18.8-18.9	80		64-67	154-156	73
72		79	31	62-63	151-153	72
71	19.0-19.3	77	30	58-61	142-150	71
70	19.4-19.5	76	29	56-57	139-141	70
69	19.6-19.7	75	28	54-55	133-138	69
68	19.8-19.9	74	27	52-53	127-132	68
67	20.0-20.1	73	26		121-126	67
66	20.2-20.3	72			115-120	66
65	20.4-20.5	71	25	50-51	112-114	65
64		70	24	46-49	109-111	64
63	20.6-20-7	69	23	44-45	106-108	63
62	20.8-20.9	68	22	42-43	103-105	62
61			21			61
60	21.0-21.1	67		40-41	100-102	60
59		66	20		97-99	59
58	21.2-21.3	65	19	38-39	94-96	58
57	21.4-21.5		18	36-37	91-93	57
56	21.6-21.7	64	17		88-90	56
55	21.8-21.9	63	16	34-35	85-87	55
54	22.0-22.1	62			82-84	54
53		61	15	32-33	78-81	53
52	22.2-22.3		14		76-77	52
51	22.4-22.5	60				51
50		59	13	30-31	73-75	50
49	22.6-22.7				70-72	49
48	22.8-22.9	58	12	28-29		48
47	23.0-23.1		11	26-27	67-69	47
46	23.2-23.3	57			64-66	46
45	23.4-23.5	56	10	24-25	61-63	45

TABLE I (Continued)

T-Score	*Obstacle race* [b]	*Pull* [c]	*Push-ups* [d]	*Sit-ups* [e]	*Bouncing* [f]	[g] *T-Score*
44	23.6-23.7	55				44
43	23.8-23.9				58-60	43
42	24.0-24.1	54	9	22-23		42
41	24.2-24.3	53			55-57	41
40	24.4-24.5		8	20-21	52-54	40
39	24.6-24.7	52	7			39
38	24.8-24.9	51			49-51	38
37	25.0-25.1	50	6	18-19		37
36	25.2-25.3				46-48	36
35	25.4-25.5	49	5	16-17	43-45	35
34	25.6-25.7	48			40-42	34
33	25.8-25.9	47	4	14-15		33
32	26.0-26.1				37-39	32
31	26.2-26.3	46	3	12-13	34-36	31
30	26.4-26.7	45			28-33	30
29	26.8-26.9	44		10-11	25-27	29
28	27.0-27.5		2			28
27	27.6-28.3	43		8-9	22-24	27
26	28.4-28.7			6-7	19-21	26
25	28.8-29.1	42	1			25
24	29.2-29.5	40		4-5	16-18	24
23	29.6-29.9					23
22	30.0-30.9	39			10-15	22
21						21
20						20
19	31.0-32.0	38		2-3	4-9	19

[a] Scales constructed on data obtained from 587 freshmen and sophomore women at University of Iowa.
[b] See page 136 for description.
[c] See page 121 for description.
[d] See page 119 for description. Subject continues as long as possible.
[e] Sit-up was taken from a lying position, legs straight and ankles held down by partner, hands were placed on shoulders with elbows close to sides. One score for each time the subject came to an erect sitting position, continuing as long as possible.
[f] See page 123 for description. Subject continues as long as possible.
[g] Construct your own score card by substituting for any of the preceding tests and inserting your own scale, and add columns for any item or items you wish to include in the battery.

it possible for a student to convert his raw score on an event into a point score. This enables him to evaluate his own performances and also provides him with a measuring stick against which to evaluate his own improvement. If it is not possible to have printed score cards, a conversion table can be posted for students' reference.

Many branches of the armed services have their achievement scales constructed for their own group since there is considerable variation in the amount of conditioning activity given different service units. Such an individual score card lends itself to the making of a profile for the student. The profile is diagnostic in that it compares abilities as measured by the tests. They consider these profiles as good motivation devices.

Table I is presented as a suggestion for an individual score card and has a profile drawn in for a student with the following record: obstacle race 21.1 seconds; pull 60 pounds; push-ups 15; sit-ups 25; bouncing 94; endurance, a T-score of 56. Observation of this profile reveals the fact that she is above average in every item except sit-ups, and that her best performance is in the agility test or obstacle race. A composite of all tests in the battery may also be included in the scale with a similar ranking on the total.

The reverse side of the card may be used for a cumulative record, other test scores or important data. If it is small and compact it is convenient for the student.

PRESENTATION OF TESTS

Before any testing is done, it is important to inform the subjects of the purpose of the test and how the results will be used. It is essential that they be interested and desirous of putting forth their best efforts. The attitude toward tests is usually a reflection on the selection of tests, the administration of them, the use of test results, or the conditioning of the students. An unfavorable attitude will exist if the tests selected have been uninteresting or meaningless, or if an undue amount of time has been taken away from the activity for testing. This will be true also if the students think that too little use has been made of the results, or if the tests have been given when they were not in physical condition for them, resulting in stiffness and soreness. Some groups will be motivated by announcing in advance that the three best scores on each team or squad will be posted. If the tests are to be used as a partial basis for marking, this should be announced, also.

The same instructions should be given to all groups. Some tests include written directions, to be read to the groups. When such direc-

tions are not available, the following principles * should be followed in preparing a set of directions.

1. The instructions should be as brief as possible, yet give an adequate understanding of what is to be done.
2. The instructions should make use of a demonstration.
3. The instructions should be adapted to the understanding of all being tested.
4. The order of instructions should be broken into units, and should be in the order of doing.
5. The instructions should equalize interest and secure maximum effort of all.

The instructions to pupils should be accompanied by instructions to examiners, mentioned earlier in the chapter under the heading "Training the Assistants." They should specify whether practice is permitted; if so, the number of trials. They should also provide for adequate and a uniform amount of rest between trials, so that fatigue will not affect the scores.

A demonstration of the test is usually advisable. The points to be emphasized should be thought out in advance and their order planned. An example of this is the Dyer wallboard tennis test, referred to further in Chapter 4.

"Demonstrate the following points: †

1. Two balls in hand.
2. Start test by dropping ball, letting it hit floor at least once, then play it.
3. Rally a few times, showing volley.
4. Cross restraining line to retrieve a ball, use a low hit to keep it in play, and retreat for the next shot.
5. Make a wild shot to show how taking another ball saves time.
6. Put this new ball in play as at the start."

All the points that apply in presenting a good demonstration for instructional purposes apply here, such as having the group placed

* Adapted from McCall, William A.: How to Measure in Education, Macmillan Company, New York, 1922, pp. 235-248.

† Quoted by permission of the *Research Quarterly*.

where they can see and hear, etc. An opportunity for questions should be given after the demonstration.

KNOWLEDGE TESTS

Care needs to be taken in administering knowledge tests as well as motor tests although the procedure is relatively simple. All too frequently teachers, through thoughtlessness or lack of information, fail to administer them in such a manner that all have a fair chance.

The room should be quiet, well ventilated, and adequately lighted. (Usually, if arrangements are made far enough in advance, physical education teachers can secure the use of a class room.) The seats should be well-spaced or students seated in alternate seats. This is of less importance when answer sheets are used, particularly those of the type shown on page 46, where copying is reduced to a minimum.

Books and wraps should not be brought into the room. A check should be made to see that all are supplied with pencils.

Answer sheets and mimeographed directions for their use can be handed the students as they enter the room. Giving test directions orally should be avoided. The directions, if general and to be used for all tests regardless of the activity, can be placed on a separate sheet from the test form. When the directions are specific to a particular test, they can be placed on the test form itself. An example of general directions to be placed on a separate sheet is included in this chapter.

Be sure that each student receives but one copy of the test questions and that all test forms are collected at the end of the examination period. If all test forms are numbered consecutively, a check can quickly be made to see that all have been returned. Students should be asked to record the number of the test form on the answer sheet. Since this enables the one in charge to locate the student who has failed to hand in the test form, it discourages the practice of carrying them away. As further insurance against allowing questions to get into circulation, require each student to personally hand in the three forms: answer sheet, directions, and test

form. Have students hand in papers when finished; this tends to avoid copying of questions.

No help should be given in the interpretation of questions, as this not only tends to give an unfair advantage to the one asking the question but also causes interruptions that are distracting.

If all are to have the same length of time, test forms can be distributed, face down, and all told to wait for a signal before starting.

The following sample of *Directions* should serve as a guide.

DIRECTIONS

You are to use an Answer Sheet to mark your answers to all of the questions in this test. Take the answer sheet now and print your name, classification, name of activity in which you are being tested, etc. Write the number of the test form in the blank space in the upper left hand corner. Then, finish reading these directions.

Exercises 1-50 in this test are of the multiple choice type, consisting of a question followed by several possible answers, only one of which is correct, or one of which is definitely better than any of the others. On the Answer Sheet you will find as many sets or rows of brackets ( ) as there are questions in the test. The number to the left of each row of brackets corresponds to the number of the question.

To answer a question, first decide which is the best answer, then find the row of brackets on the Answer Sheet numbered the same as the question. Then mark a cross (X) in the brackets corresponding to the correct or best response, COUNTING FROM THE LEFT. If the first response is correct or best, place a cross in the first brackets in the set; if the second response is correct or best, place a cross in the second brackets, etc.

Exercises 51-75 are of the true-false type. If the statement is entirely true, place an (X) in the first brackets; if partially or entirely false, place the cross in the second set of brackets. All omissions will be counted as errors. The sample questions have been marked correctly on the Answer Sheet.

Answer the questions in the order in which they are given, but do not linger too long over difficult items. Skip those and return later if time permits. If you do skip an exercise, be sure to skip the corresponding row of brackets on the Answer Sheet. Any mark which you unintentionally place in the wrong place will count against you. If you change your mind, thoroughly erase your first mark. NEVER PLACE MORE THAN ONE CROSS IN ANY ROW.

Do not begin work until you are told to do so. If you have any questions, ask them now.

Answer sheets, commonly used in wide scale testing in academic subject matter, have so many advantages that they should be in more common usage in physical education. A few advantages are listed below:

1. Save paper, stencils, and secretarial time as the test forms can be used repeatedly.
2. Can be more accurately and conveniently scored than thumbing through pages of questions for each student.
3. Save teacher time.

A master answer sheet, that can be used for either multiple choice of true-false questions in any test where the total number of questions does not exceed seventy-five, is shown in the illustration. This saves preparing answer sheets for each examination. The sheet can be made to accommodate more questions by reducing the size of the type.

—— Number of Test Form

ANSWER SHEET

Please do not write in space below.

NAME _____
(*last name*) (*first name*)

Classification _____ Date _____, 19_____

Activity _____ Instructors Name _____

Sample for multiple choice:

0. ( ) ( ) (X) ( ) ( )

	1	2	3	4	5
1.	( )	( )	( )	( )	( )
2.	( )	( )	( )	( )	( )
3.	( )	( )	( )	( )	( )
.					
.					
.					
25.	( )	( )	( )	( )	( )
	1	2	3	4	5

	1	2	3	4	5
26.	( )	( )	( )	( )	( )
27.	( )	( )	( )	( )	( )
28.	( )	( )	( )	( )	( )
.					
.					
.					
50.	( )	( )	( )	( )	( )
	1	2	3	4	5

Sample for true-false:

0. (X) ( ) ( ) ( ) ( )

	1	2	3	4	5
51.	( )	( )	( )	( )	( )
52.	( )	( )	( )	( )	( )
53.	( )	( )	( )	( )	( )
.					
.					
.					
75.	( )	( )	( )	( )	( )
	1	2	3	4	5

Figure 9. Sample of Answer Sheet for Knowledge Tests

Punched keys for use in scoring answer sheets can be quickly and easily made by following the directions listed below:

1. Attach a blank answer sheet to a piece of lightweight cardboard with paper clips, inserting a piece of carbon paper between the two. (Manila filing folders work very well.)

2. Label the key with name of the examination.
3. Trace a few of the question numbers, to be used later as a guide when superimposing the key on the answer sheets.
4. Place crosses in all of the appropriate brackets, indicating correct answers.
5. Remove the dummy answer sheet and carbon paper and proceed to use a paper punch on all of the crosses. Paper punches with wide jaws are best for this purpose, since they permit reaching across to the center of the cardboard without rolling or folding. An ordinary "dime store" paper punch can be used if you cut the cardboard lengthwise into thirds. After completing the punching process, the cardboard can be taped together again.
6. Check to see that the holes have been properly placed by placing the key on the dummy of correct answers.

In scoring, superimpose the cardboard key on the answer sheet, being careful to place it in proper position. If the margins of the key are trimmed so that the question numbers appear, placement is facilitated. Place a red dot in any hole where a cross does not appear, counting the number of errors and record the number on the answer sheet. This is of convenience in going over examinations later with students as it locates their errors and also indicates the right answer for each question they missed.

BIBLIOGRAPHY

1. Brophy, Kathleen: "Target for Testing Accuracy in Softball Throwing," *Softball-Volleyball Guide,* A. S. Barnes & Company. New York, 1939, p. 29
2. Schmithals, Margaret, and French, Esther: "Achievement Tests in Field Hockey for College Women," *Research Quarterly,* 11, October 1940, p. 840
3. Scott, M. Gladys: "Achievement Examinations in Badminton," *Research Quarterly,* 12, May 1941, p. 242
4. Wetzel, Norman C.: "The Simultaneous Screening and Assessment of School Children," *Journal of Health and Physical Education,* 13, December, 1942, p. 576.
5. Women's Advisory Committee on Physical Education, California State Superintendent of Schools: "Physical Education—A Wartime and Peacetime Program for Girls," Stanford University Press, 1942, pp. 225-228.

4.

Measurement of Skill in Sports

The development of skill is widely recognized as one of the major objectives in physical education. Measuring sports skills is an important aspect of the teaching procedure. Skill tests have been used for a considerable length of time, but many of the earlier tests had no statistical work done on them. Therefore, their reliability and validity were questionable. Also, it was not known how much they duplicated one another, or the extent to which they measured general motor ability rather than ability in the activity *per se*. Considerable experimentation has been done in the last decade, and some of the results have been published. The reports of these studies are not readily available to teachers, since they are scattered throughout the physical education literature, and some have not been published, heretofore. We will attempt in this chapter to present the ones which are of proven value and to include enough information on each that they can be used. Tentative standards, when available, are suggested for many of the tests. The source of the data for each scale is given so that the reader will have some basis for deciding whether the scale is applicable to the class for which the test is being used.

It is not our purpose to duplicate the work of previous writers in presenting a critical review of all published tests.

BASIS FOR SELECTION

Tests have been selected which come closest to meeting the criteria set forth in Chapter 2, Characteristics of A Good Test. The bibliography presented at the end of this chapter is provided for those persons wishing to study more of the available published tests, or those persons wishing to give a more extensive battery of tests.

STANDARDS OR ACHIEVEMENT SCALES

The standards or achievement scales presented here are based on scores made by fairly limited and sometimes rather select groups. The group is described in each case. The scales are not to be interpreted in the same manner as nation-wide norms. Unfortunately, data have not been collected for boys and men, though the majority of the tests should be useful for that group. This is not too great a handicap as usable scales can be constructed on scores made by classes in any school. Frequently this is necessary even though other scales are available.

BADMINTON

1. SERVE

Equipment: A clothesline rope stretched 20 inches directly above the net and parallel to it, attached to the same standards as the net. New birds and tightly strung racquets.

2. Floor markings

Using the intersection of the short service line and the center line as a midpoint, describe a series of arcs in the right service court at distances of 22 inches, 30 inches, 38 inches, and 46 inches from the midpoint, measurement including the width of the 2 inch line. Extend these arcs from the short service line to the center line, as indicated in the diagram. (Figure 10) The lines should be painted in different colors to increase accuracy in scoring. Show card paint, which can be washed from the floor, is suggested.

Test: The player being tested shall stand any place in the service area diagonally opposite the target, and shall serve twenty birds, attempting to send them through the space between the rope and the net in such a manner that they land in the right service court for the doubles

game. The scorer shall stand near the center of the left service court on the same side of the net with the target and facing the target. The corner of the target nearest the intersection of the service line and center line counts five points, the next space four points, the next three, then two, and any bird off the target but in the service area for the doubles game counts one point.

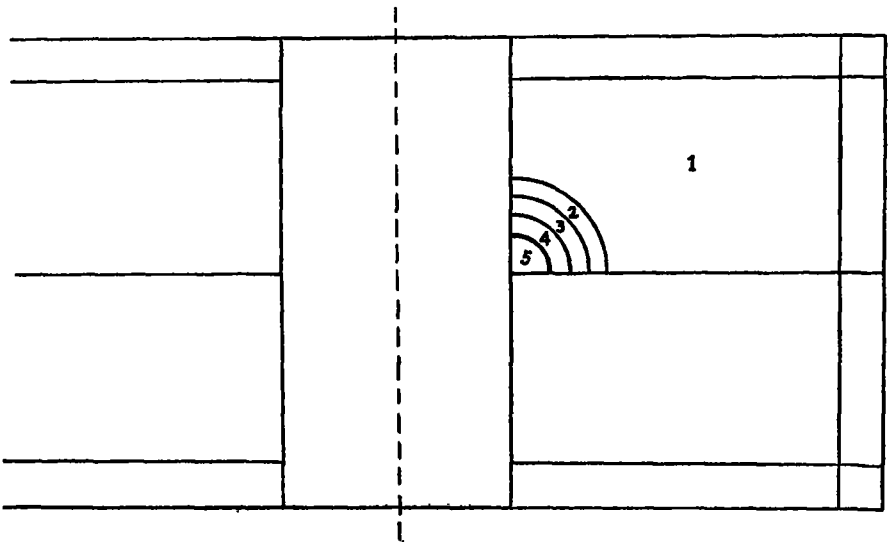


Figure 10. Floor Markings for Badminton Serve Test

1-5 score for respective areas

Scoring: No score for any trial which fails to go between the rope and the net or which fails to land in the service court for the doubles game. Any bird landing within an area or on the line surrounding an area is scored as shown in the diagram. Any bird landing on a line dividing two scoring areas shall receive the score of the higher area. The score for the entire test is the total of twenty trials. It is considered a foul and the trial is repeated if the serve is illegal.*

* For definition of legal serve, see American Badminton Association rules.

Reliability: Odd-even method, stepped up by the Spearman-Brown Prophecy Formula, .88, computed on 29 major students, State University of Iowa.*

Validity: The validity of the test was found to be .66, when correlated with a criterion of tournament standings (ladder tournament carried on throughout twenty class periods).

T-scores: Computed on over two hundred University of Minnesota students (not physical education majors) at the end of about thirty class periods of instruction. See page 55.

Reference: Scott, M. Gladys: Achievement Examinations in Badminton, *Research Quarterly,* 12, May 1941, p. 242. (Quoted by permission of the *Research Quarterly.*)

Comments: This test measures accuracy of placement and also the ability to serve the bird in a low flight. It is easy to administer, and can be given off the courts, so that it does not interfere with play. See Chapter 3 for suggestions. The amount of practice should be held constant for all players and the test should not be administered until the majority have acquired the ability to make short, low serves. The condition of equipment affects the scores decidedly.

2. CLEAR

Equipment: 1. A clothesline rope stretched across the court 14 feet from the net and parallel to it, at a height of 8 feet from the floor.

2. Floor markings

a. Construct a line 2 feet nearer the net than the rear service line in the doubles game and parallel to it. Measure from the exact center of the line. Extend this line

* See p. 241 for a discussion of the statistical procedures.

from one outer alley line to the other outer alley line.

b. On the same side of the net, construct a line 2 feet farther from the net than the rear service line in the single game and parallel to it. Measure from the exact center of the line. Extend this line from one outer alley line to the other outer line. The lines should be painted different colors to increase accuracy in scoring.

c. On the opposite side of the net, draw marks 2 inches square at spots indicated on the diagram as X and Y. The center of X should be 11 feet from the net and 3 feet from the center line toward the left sideline. The center of Y should be 11 feet from the net and 3 feet from the center line toward the right sideline. In measuring from the center line, use the exact center of the line.

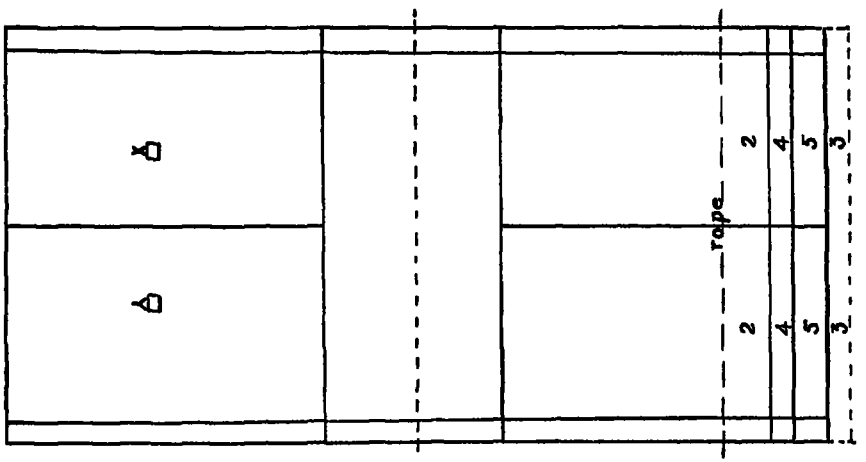


Figure 11. Floor —Markings for Badminton Clear Test

2-5 score for respective areas X-Y limits of set up for clear stroke

Test: The player being tested shall stand between the two square marks, on the court opposite the target. The person giving the tests (player with

considerable experience) shall stand on the intersection of the short service line and the center line on the same side of the net as the target and shall serve the bird to the player being tested. The bird must cross the net with enough force to carry it to the line between the two squares before it touches the floor. (This is an imaginary line.) If it does not go that far or is outside the space between the two squares, the player being tested should not play it. The player being tested may move any place he wishes as soon as the bird has been hit to him. Only birds played by the player being tested shall count as trials. He shall attempt to send the bird by means of a clear stroke above the rope so that the bird lands on the target. Twenty trials are allowed. The person giving the test should call out the score of each trial, to be recorded by an assistant. The area between the two rear lines of the regulation court counts five points, the space just behind it counts three points, and the space just in front of the two rear lines of the regulation court counts four points. Any bird going over the rope but failing to reach the target counts two points.

This test can be given to two players at once on the same court, extending the imaginary line farther.

Scoring: No score for any trial failing to go over the rope or failing to land in the court in the space behind the rope and on the target, as indicated on the diagram. Any bird landing within an area or on the line surrounding the area is scored as shown in the diagram. Any bird landing on a line dividing two scoring areas shall receive the score of the higher area.

The score for the entire test is the total of twenty trials. It is considered a foul and the trial is repeated if the stroke is "carried" or "slung." *

Reliability: Correlation by the odd-even method, .91; stepped up with the Spearman-Brown formula, .96. The subjects were the same as for Test 1.

Validity: .60 with a criterion of tournament rankings.

T-scores: See page 55.

Reference: Same as for Test 1.

Comments: This test, even with the disadvantage of the bird being put in play by another player, is so highly reliable that it would appear that experimentation should be done to determine the effect of fewer trials. The comparatively low validity also indicates the need for further study. It is an excellent practice device. It measures power and to some extent accuracy in the strokes. It would also seem logical that it is a measure of the player's judgment.

BATTERY OF BADMINTON TESTS

The one best test appears to be the serve test, as described here. The experimental battery included six tests, two each on serve, clear, and drop. The validities on the drop tests were low. The clear and serve tests described here have a low intercorrelation, .11, and therefore, appear to be measuring quite different abilities. When the scores on the two tests are combined, by means of multiple correlation, the validity coefficient was .85, which indicates that they make a satisfactory battery. The formula for insuring proper weighting to the two tests, when combined, is as follows:

$$1.\ \text{serve} + 1.2\ \text{clear}$$

* See official A.B.A. rules for interpretation of terms.

T-SCORES FOR BADMINTON TESTS

These scales were constructed from data collected in college classes, University of Minnesota. The tests were administered at the end of about thirty class periods of instruction, each period thirty-five minutes in length. This was the first season of badminton for the great majority.

TABLE II

T-SCALES FOR BADMINTON TESTS

SERVE TEST				CLEAR TEST			
Raw Score	*T-Score*	*Raw Score*	*T-Score*	*Raw Score*	*T-Score*	*Raw Score*	*T-Score*
70 up	79	40	51	94-96	78	48-49	44
69	74	39	50	92-93	74	46-47	43
68	72	38	49	90-91	72	44-45	42
67	71	37	48	88-89	70	42-43	42
66	69	36	48	86-87	67	40-41	41
65	69	35	47	84-85	64	38-39	41
64	68	34	47	82-83	61	36-37	40
63	67	33	46	80-81	59	34-35	39
62	66	32	45	78-79	57	32-33	39
61	65	31	45	76-77	56	30-31	38
				74-75	55	28-29	38
60	63	30	44	72-73	54	26-27	37
59	62	29	44	70-71	52	24-25	36
58	62	28	43	68-69	51	22-23	36
57	61	27	43	66-67	50	20-21	35
56	60	26	42	64-65	49	18-19	34
55	59	25	41	62-63	49	16-17	32
54	59	24	41	60-61	48	14-15	32
53	58	23	40	58-59	47	12-13	32
52	58	22	40	56-57	46	10-11	31
51	57	21	39	54-55	46	6- 9	30
				52-53	45	2- 5	27
50	56	20	39	50-51	45		
49	56	19	38				
48	55	18	38				
47	55	17	36				
46	54	16	35				
45	53	15	34				
44	53	14	33				
43	53	13	32				
42	52	12	31				
41	51	11	31				
		10	29				
		5-9	27				
		1-4	23				

BADMINTON SCALES FOR PHYSICAL EDUCATION MAJOR STUDENTS

The scales below are indicative of what may be expected of physical education majors at the end of seven lessons, 35 minutes in length. The subjects were 43 women, majors in physical education, University of Minnesota. The scores are translated into a grading plan: 7% A, 24% B, 38% C, 24% D, 7% failure.

SERVE TEST		CLEAR TEST	
Range: 13-70		*Range: 6-86*	
56 and up	A	81 and up	A
45-55	B	72-80	B
25-44	C	56-71	C
20-24	D	34-55	D
19 and below	Fd.	33 and below	Fd.

BASKETBALL

1. HALF-MINUTE SHOOTING

Equipment: No special equipment. Balls should be well inflated. Stop watch.

Test: Player stands at any position he selects near the basket, with a ball in his hands. On the signal, Ready, Go! he starts shooting and continues until the signal to stop, attempting to make as many baskets as possible within the 30 seconds. If the ball has left his hands when the signal to stop sounds, the basket counts, if made. Two trials are given each player.

Scoring: The number of baskets made in 30 seconds is the score for each trial. The better of the two trials is recorded.

Reliability: $r = .58$ on trials 1 and 3 versus trial 2; stepped up by the Spearman-Brown formula, .68, for 107 girls in Proviso Township High School, Maywood, Illinois. (Jones)

$r = .73$ on 190 high school boys. (Johnson)

r = .54 on first and second trials; subjects were 209 college freshman women, University of Iowa. (Scott)

Validity: r = .58 for the sum of three trials with a rating criterion. (Jones)
.71 on 190 high school boys. (Johnson)
.60 with a sports tests criterion, 155 college freshman women, University of Iowa. (Scott)

T-scores: See page 58.

References:
1. Johnson, L. W.: *Objective Tests in Basketball for Boys.* Unpublished M. A. thesis, University of Iowa, 1934.
2. Jones, Edith: *A Study of Knowledge and Playing Ability in Basketball for High School Girls.* Unpublished M. A. thesis, University of Iowa, 1941.
3. Scott, M. Gladys: "Assessment of Motor Abilities of College Women through Objective Tests," *Research Quarterly,* 10. Oct., 1939, p. 63.

Comments: This test measures the ability to hit the spot at which one is aiming, and also the ability to judge rebounds, to move quickly to get to the ball and to put it in play again quickly. This makes it a good test for all players, regardless of position. If there are as many as six baskets available, the test can be administered in a very few minutes to the usual class. Since no two backboards are exactly alike, the conditions for all can be somewhat equalized by having players rotate to a new basket for the second trial. A minimum of three trials is recommended for relatively inexperienced players.

2. WALL PASS

This test is described in Chapter 6, Measurement of General Motor Ability.

TABLE III

T-Scales for Basketball Tests

T-Score	Half Minute [a] Basket Shooting (276) [d]	Half Minute [b] Basket Shooting (486) [d]	Wall Pass [c] (105) [d]	T-Score	T-Score	Half Minute [a] Basket Shooting (276) [d]	Half Minute [b] Basket Shooting (486) [d]	Wall Pass [c] (105) [d]	T-Score
81		18		81	50			34	50
80				80	49	5	5		49
79	15			79	48				48
77		17		77	47				47
76		16	42	76	46		4	33	46
75	14			75	45				45
74		15		74	44				44
73				73	43	4		32	43
72	13	14		72	42				42
71	12		41	71	41		3	31	41
70		13		70	40				40
69				69	39				39
68	11	12	40	68	38				38
67				67	37	3		30	37
66	10	11		66	36		2		36
65			39	65	35				35
64	9	10		64	34				34
63				63	33			29	33
62		9		62	32				32
61	8		38	61	31		1		31
60				60	30			28	30
59		8		59	29	2			29
58	7		37	58	28			27	28
57				57	27				27
56		7	36	56	26				26
55				55	25				25
54	6			54	24			26	24
53		6	35	53	23				23
52				52	22				22
51				51	21	1			21

[a] Scale was constructed on data obtained from girls in ninth grade in Blue Island Community High School, Blue Island, Illinois, and eleventh grade in Proviso Township High School, Maywood, Illinois. The scores were on the best one out of two trials.

[b] Scale was constructed on data obtained from college students in University of Iowa and University of Minnesota.

[c] Scale was constructed on data obtained from girls in eleventh grade in the Proviso Township High School, Maywood, Illinois.

[d] Number of cases in the distribution on which the scale was constructed.

Reliability: r = .59 on trials 1 and 3 versus 2, with Spearman-Brown formula, .68. (Jones)
r = .62 on a single trial on two successive days, 188 college freshman women, University of Iowa. (Scott)

Validity: r = .44 for the sum of three trials with a rating criterion. (Jones)
r = .62 with a sports test criterion, 155 college freshman women, University of Iowa.

T-scores: See Tables XI and XII.

References: See references for Test 1. (Jones, Scott)

Comments: This test measures the ability to pass and catch rapidly. It can be administered at the same time that the half-minute basket shooting test is being given, if wall space permits. More trials may be helpful.

BATTERY OF BASKETBALL TESTS

It is not known how much the two tests given here overlap. If time permits the administration of just one test, the half-minute shooting test should be given. If only two baskets are available, the half-minute test will take considerable time to give. In that case only part of the class should be assigned to tests at a given time, and the rest of the class can practise on other techniques in the center of the floor. If the wall pass has been given in the general motor ability battery, it need not be repeated at the beginning of the basketball season.

FIELD HOCKEY

1. Dribble, Dodge, Circular Tackle, and Drive

Equipment: 1. Hockey stick for each participant; stop watch; one or two balls; and two high jump standards (or other portable post) with round bases.

2. Field markings
   (a) A line 20 feet long to be used for a starting line.
   (b) A line perpendicular to the midpoint of the starting line and extending 35 feet from it. This is the foul line.
   (c) A line 10 feet long, perpendicular to and bisected by the foul line at a point 30

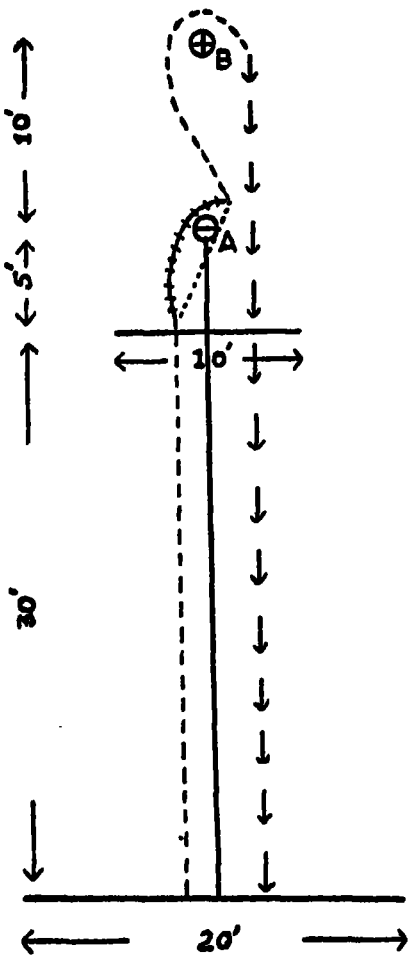


Figure 12. Field Markings and Action Sequence for Field Hockey Test 1

A-B jump standards
— — dribble
..... path of ball (in dodge)
++++++ path of player (in dodge)
→→→ drive

feet from the starting line. This is the restraining line.

(d) A line 1 foot long, perpendicular to and bisected by the foul line at a point 35 feet from the starting line.

(e) Two lines, each 1 foot long, bisecting each other at a point which is 45 feet from the starting line and in a straight line with the foul line. See Figure 12.

3. Position of standards

(a) One standard is placed so that the middle of the base of the standard is directly over the point where the foul line and the line described in 2(d) bisect each other.

(b) The other standard is placed in a similar fashion over the point formed by the two lines described in 2(e).

Test: The player being tested shall stand behind the starting line with the hockey ball placed on the starting line at any point to the left of the foul line. At the signal, Ready, Go! the player shall dribble the ball forward to the left of and parallel to the foul line. As soon as the restraining line is reached, the ball shall be sent from the left side of the foul line to the right of the first obstacle (from the player's point of view), and the player shall run around the left side of the obstacle and recover the ball. (This is analogous to the dodge.) Next, the player shall execute a turn toward her right around the second obstacle, still keeping control of the ball. (This is analogous to the circular tackle.) As soon as possible after that the ball shall be driven toward the starting line. If the drive is not hard enough to reach the starting line the player must follow it up and hit the ball

again. This procedure shall be repeated until six trials have been given; players are alternated on trials to avoid their becoming fatigued.

Scoring: The score for one trial shall be the time it takes from the signal "Go" until the player's ball has again crossed the starting line. The score for the entire test is the average of the six trials. It is considered a foul and the trial does not count if:

1. The ball or player crosses the foul line before reaching the restraining line.
2. In executing the dodge, the ball is not sent from the left side of the foul line.

Reliability: r = .92 when computed by the odd-even method, and stepped up by the Spearman-Brown formula. The subjects were 51 players in college classes and hockey club, University of Iowa.

Validity: r = .44 when correlated with criteria determined by ratings made by three nationally rated umpires.

T-scores: See page 69.

Reference: Schmithals, Margaret and French, Esther: "Achievement Tests in Field Hockey for College Women," *Research Quarterly,* 11, October, 1940, p. 84. (Quoted by permission of the *Research Quarterly.*)

Comments: This test has been used for several years, in combination with a test of knowledge, for early season classification, and serves adequately. It calls for the use of a variety of skills, the ability to control the ball while moving rapidly, and to make quick changes of direction. It can be administered along the edges of the field and requires no special equipment, other than a stop watch and jumping standards. The test authors think that the low

validity may be partially due to the difficulty in making discriminations in ratings. Field hockey presents a very difficult rating problem, with so many players and the small number of times that the ball is played by some players. The judges rated the players on two successive playing periods, which was not adequate to secure a wide range in ratings.

2. GOAL SHOOTING—STRAIGHT, RIGHT, LEFT

Equipment: 1. Target, 9 inches wide, 12 feet long and at least ½ inch thick, made of hard wood. The board is painted according to the following specifications: The length of the board is divided into eleven equal spaces, alternate space starting from either end being painted black and the other remaining the natural color of the wood. Numbers are painted in

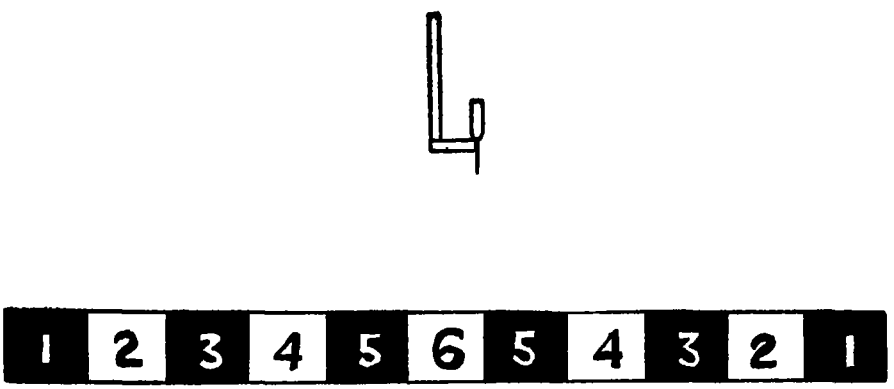


Figure 13. Target for Hockey Test 2

a – side view
b – front view and scoring scheme

the spaces in contrasting colors (black on light background and white on black background) in the following order starting from either end: 1-2-3-4-5-6-5-4-3-2-1 (See Figure 13). A base made of board at least 3 inches

wide, exactly 12 feet long, and at least ½ inch thick, is nailed on the bottom of the target so that 2½ inches extend beyond the back of the target. The board, in order to stand upright securely, may be anchored with an ice pick or other similar device.
Hockey stick for each participant, four to ten balls, stop watch.

2. Field markings
   (a) A line 6½ feet long to be used as a starting line.
   (b) A rectangle, 11 feet long and 6½ feet wide, 15 feet from the starting line. Point A is the midpoint of the side opposite the starting line.
   (c) A line 12 feet long, called the center target line, parallel to and 60 feet from the starting line.
   (d) A line 12 feet long, called the right inner target line.
   (e) A line 12 feet long, called the left inner target line.
   (f) The target is placed directly on the specified line with the numbers facing the starting line and the board anchored with ice picks. For the straight drive, it is placed on the center target line, for the drive from right and left inners' positions, the right and left inners' target lines, respectively. (See Figure 14)

Test: 1. Drive from center's position. The player being tested shall stand behind the starting line with the hockey ball placed on the starting line. At the signal, Ready, Go! the ball shall be dribbled to the rectangle, from within which area it must be driven toward the board (placed on the center target line). The

procedure shall be repeated until ten trials have been given.

2. Drive from right inner's position. The same procedure shall be repeated, the only difference being the position of the target, which is placed on the right inner target line.
3. Drive from left inner's position. The same procedure shall be repeated, the only difference being the change in position of the target to the left inner target line.

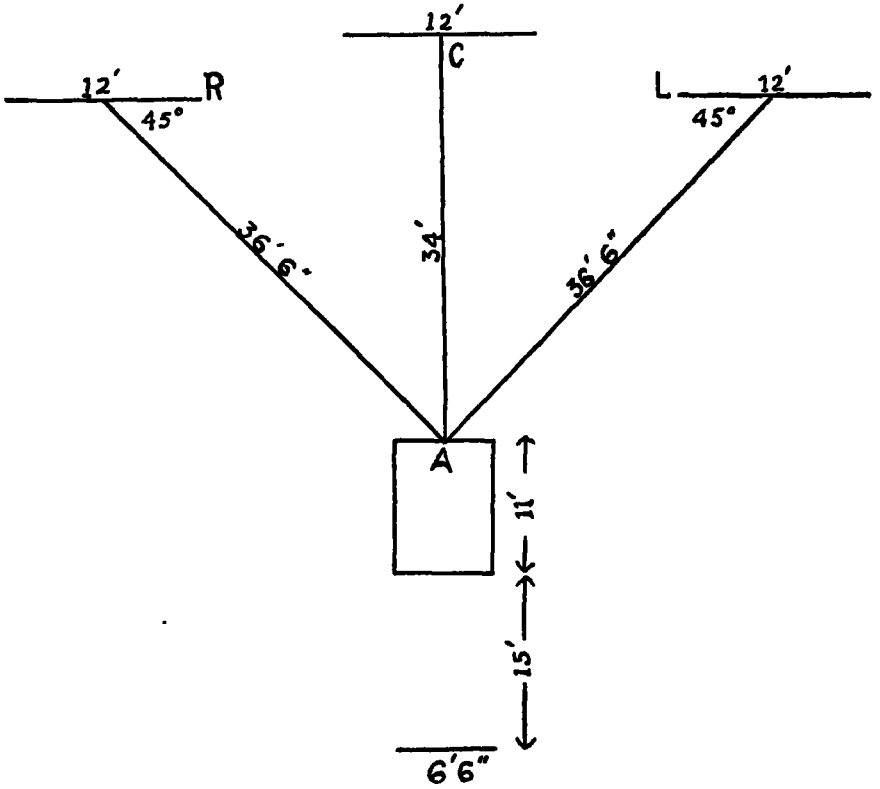


Figure 14. Field Markings for Hockey Test 2

A — center of front line of rectangle, from which target lines are measured
R — right inner target line
C — center target line
L — left inner target line

Scoring: The score for one trial shall constitute the time elapsing from the timer's signal, Go! until the ball strikes the target. The score for the

entire test is the sum of the first and second best odd and first and second best even numbered trials made on the center drive, the right inner drive, and the left inner drive. The score shall count if the ball bounces over the top of the target; in this case, the time shall be taken until the instant the ball clears the target. The score shall be zero if the ball is not driven from within the rectangle, or if the driven ball fails to reach the target or misses it at either end. The attempt shall not be counted as a trial if "sticks" are made, or if the player raises the ball so that it fails to touch the ground before passing above the target.

Reliability: r = .92 when the Spearman-Brown Prophecy formula is applied. The subjects are the same as for Test 1.

Validity: r = .48 with the same criterion and for the same group as Test 1.

T-scores: See page 69.

Reference: Same as for Test 1.

Comments: This test is designed to measure the ability of the player to adjust footwork, to judge space while moving, and to drive with accuracy and force while running. It is an excellent practice device, and can be divided into parts, administering just one (goal shooting left), if time is limited. The reliability on goal shooting left is adequate (.87) and the validity is substantially the same as for the entire test (.44). Note that the accuracy score is used to make the situation somewhat analogous to the game when the player has to decide how much speed she can afford to sacrifice for the sake of accuracy. The player is given two scores for this test: a speed score, and an accuracy score. The time score

is the only one that actually is used in combining this test with others in a battery.

3. FIELDING AND DRIVE

Equipment: 1. Hockey stick for each participant, three to seven balls, two ice picks with brightly colored tops, regulation hockey goal, stop watch.

2. Field markings (see Figure 15)

(a) Goal line extending across the area between goal posts.

(b) Foul line, 12 feet long, parallel to and 10 feet from the goal line, located directly in front of the goal area. The ice picks are placed on the foul line at points directly opposite each goal post.

(c) Restraining line, 30 feet long, parallel to and 10 feet from the foul line.

(d) Regulation striking circle in front of the goal.

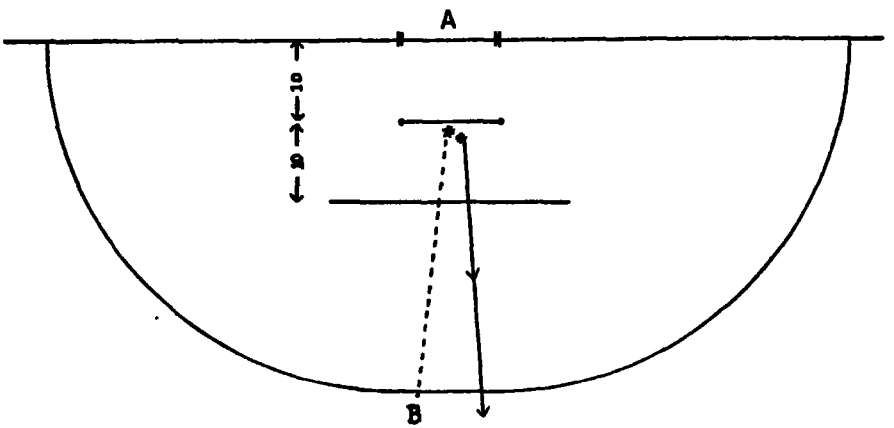


Figure 15. Field Markings for Hockey Test 3

A player in position to start the test
B examiner in position to roll ball
– – – rolled ball
* stop
≡ tap
→ drive

Test: The player being tested shall stand behind the goal line. The examiner shall stand at the edge of the striking circle directly in front of the goal

with a hockey ball in one hand and a stop watch in the other. At the examiner's signal, Ready, Go! the hockey ball is rolled toward the goal. Simultaneously, the player shall run forward and attempt to field the ball before it reaches the foul line, tap it once, and drive it out of the striking circle from within the area between the restraining line and the foul line. This procedure shall be repeated until sixteen trials have been given. Instructions to examiner: The ball should be rolled as uniformly as possible for all trials and for all players. Through a little practice a roll may be achieved which is approximately 45 feet in 1.7 seconds.

Scoring: The score for one trial is the time from the instant the player first touches the hockey ball to the instant the ball reaches the edge of the striking circle. The score on the entire test is the sum of the average of the three best even and the three best odd numbered scores of the sixteen trials. The attempt does not count as a trial if the rolled ball does not pass between the two ice picks, or if not delivered at approximately the designated speed. The player receives a zero score on a trial if the ball is advanced illegally, or if it rolls wholly over the foul line before or after it is touched by the player's stick. The zero score is assigned, too, if the ball is not driven out of the striking circle from within the area bounded by the restraining line and the foul line, or if the ball is not controlled; that is, stopped, and tapped, before being driven.

Reliability: $r = .90$ by the odd-even method, stepped up by the Spearman-Brown formula. The subjects were the same as for Test 1.

Validity: r = .48 with the same criterion as described for Test 1.

T-scores: See below.

Reference: Same as for Test 1.

Comments: The effect on the score of having the ball put in play by another person than the one being tested appears to be very slight, as the procedure is well standardized. This test is costly in time to administer.

BATTERY OF FIELD HOCKEY TESTS

The best single test for classification of college students is the dribble, dodge, circular tackle, and drive. The best combination of two tests, as determined by multiple correlations, is the goal shooting left (one portion of test two) and the fielding and drive test. They measure quite different things as is indicated by the low intercorrelation of .22. The two tests together yielded a correlation coefficient of .60. The addition of Test 1 to this battery raises the correlation only slightly (.62). The formula for combining scores on the two tests (goal shooting left and the fielding and drive) to insure proper weighting is as follows:

1. goal shooting left + 1.2 fielding and drive.

TABLE IV

T-SCALES FOR FIELD HOCKEY TESTS

T-Score	*Dribble, Dodge,*[a] *Circular Tackle, and Drive*	*Goal*[b] *Shooting*	*Fielding*[b] *and Drive*	*T-Score*
73		19.4-19.6	4.1	73
71	10.1-10.5	19.7-21.4	4.3	71
69				69
68		21.5-21.7	4.4	68
67	10.6-11.0			67
66				66
65	11.1-11.5		4.5	65
64		21.8-22.0	4.6	64
63				63
62			4.7	62
61	11.6-12.0	22.1-22.3	4.8	61

TABLE IV (Continued)

T-SCALES FOR FIELD HOCKEY TESTS

T-Score	*Dribble, Dodge,[a] Circular Tackle, and Drive*	*Goal Shooting*	*Fielding and Drive*	*T-Score*
60		22.4-22.6		60
59		22.7-22.9	4.9	59
58	12.1-12.5	23.0-23.2		58
57		23.3-23.8	5.0	57
56	12.6-13.0	23.9-24.1	5.1	56
55		24.2-25.0	5.2	55
54	13.1-13.5	25.1-25.3	5.4	54
53		25.4-25.6	5.6	53
52	13.6-14.0		5.8	52
51		25.7-25.9	5.9	51
50	14.1-14.5	26.0-26.5	6.0	50
49		26.6-26.8	6.1	49
48	14.6-15.0	26.9-27.1	6.2	48
47	15.1-15.5	27.2-27.4	6.4	47
46		27.5-28.0	6.5	46
45	15.6-16.0	28.1-28.3		45
44				44
43	16.1-16.5	28.4-28.6	6.6	43
42	16.6-17.0	28.7-28.9		42
41		29.0-29.5	6.7	41
40	17.1-17.5	29.6-29.8	6.9	40
39		29.9-30.7	7.2	39
38	17.6-18.0	30.8-31.0	7.3	38
37	18.1-18.5	31.1-31.9		37
36	18.6-19.0	32.0-32.8	7.4	36
35			7.6	35
34	19.1-19.5	32.9-33.1		34
33	19.6-20.0		7.7	33
32		33.2-33.4	8.0	32
31	20.1-20.5	33.5-33.7	9.0	31
30				30
29	20.6-21.0	33.8-34.3	9.2	29
28				28
27	21.1-24.5	34.4-34.6	10.7	27
25	24.6-30.0			25

[a] This scale was constructed on data collected on 81 students, Illinois State Normal University. Average of three trials was used.

[b] These scales were constructed on data collected on 51 students, University of Iowa.

SOCCER

1. VOLLEYING

Equipment: 1. One soccer ball, fully inflated; one stop watch; one unobstructed wall space 15 feet long and 10 or more feet high.

2. Markings

(a) A target shall be outlined on the wall which is 15 feet wide, beginning 8 inches above the floor and continuing upward to at least 10 feet from the floor.

(b) A starting line, 2 feet in length, shall be drawn on the floor, parallel to the wall and 15 feet from the midpoint of the outlined wall target.

(c) An area shall be outlined on the floor which is 30 feet square, and it shall have as one side of the square the wall space upon which the target is marked, so that their midpoints coincide. See Figure 16.

Test: Place the ball on the starting line. At the signal, Ready, Go! kick the ball against the wall so that it hits within the outlined target; when the ball rebounds, recover it and kick again. Continue this as rapidly as possible. Your score is the number of times the ball strikes the wall within the target in one minute. You may kick, volley, or use any technique which is legal in a regulation soccer game. After the first kick, the ball may be played from the point of recovery (provided that this point is within the 30-foot square on the floor) or you may dribble the ball to a more advantageous position. If the ball bounds outside the 30-foot square, it will be stopped by one of the assistants and placed on the 30-foot square boundary line, where you may recover it and dribble or kick it again.

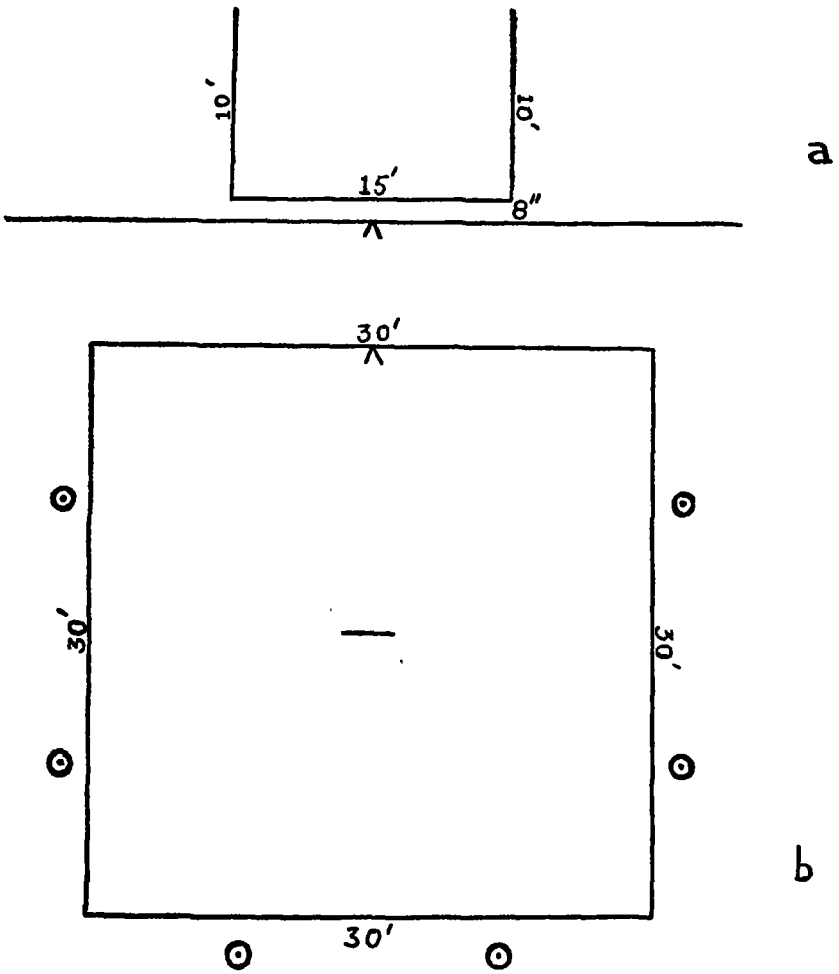


Figure 16. Wall and Floor Markings for Soccer Volleying Tests

a — wall target
b — floor diagram
∠ — center of floor area and wall area (the two should coincide)
⊙ — assistants to recover balls going out of the area

Note to the test administrator: Station six assistants, two on each of the free sides of the 30-foot square as marked on the diagram by ⊙'s. These assistants are to stop any ball which is about to leave the square and place it, with their hands, on the boundary line at the point where it crossed the line.

Practice: Allow each player one practice trial of one-half minute length.

Scoring: Two trials shall be given each player, each on a separate day. Not more than a week should elapse between trials. The better of the two trials shall count. One point is scored each time the ball strikes within the correct area. Balls hitting lines outlining the target are considered good.

Reliability: The coefficient was .67 for first and second trials. The subjects were 84 ninth and tenth grade girls, Fairview High School, Fairview Village, Ohio. The tests were given at the end of the season.

Validity: The validity coefficient of this test was found to be .57 when correlated with the criterion of subjective ratings, established by a board of judges consisting of three instructors of soccer and two senior students. Another criterion was established by T-scoring the tests and adding the T-scores made on all tests. When correlated with this criterion, the coefficient of the volleying test was .77. When such a criterion is used the coefficient is always higher because the test being studied is a part of the total score used for the criterion.

T-scores: See page 83.

Reference: Schaufele, Evelyn F.: The Establishment of Objective Tests for Girls of the Ninth and Tenth Grades to Determine Soccer Ability, Unpublished M.A. thesis, University of Iowa, 1940.

Comments: The test appears to measure ability to control the ball, judgment of speed and direction, and skill in maneuvering the ball, all of which are essential skills in the game itself. It is easy to administer and is interesting to students. An increase in the number of trials

appears to be indicated. Four should be sufficient.

2. PASSING AND RECEIVING

Equipment: 1. One soccer ball, fully inflated; one unobstructed wall space 55 feet in length; one stop watch.

2. Markings

(a) A restraining line shall be drawn 8 feet from the wall and parallel to it.

(b) Two boundary lines, perpendicular to the wall and 12 feet in length, shall be drawn at the end of each restraining line, extending out from the wall. These lines shall be 55 feet apart. See Figure 17.

Test: Place the ball at the point where the restraining line crosses the boundary line, with the wall on your left (indicated by A in the diagram). On the signal, Go! dribble the ball forward a few feet and then pass it with the side of the foot so that it strikes the wall and rebounds. Run forward to meet it as it rebounds, and repeat. You should wait for the ball to cross the restraining line, as you may not make a pass from inside the area. However, if the ball has not been sent with enough force, you may go in to recover it and dribble it outside the restraining line. You have ten seconds to complete the distance, during which time you are to make three passes against the wall, if possible. Your score will be the number of successful passes *and recoveries* you make. Your third pass and recovery will not count unless you have touched the ball beyond the finish line before the time limit. If you recover the ball before you reach the finish line, try to dribble it up to the line before time is up. The

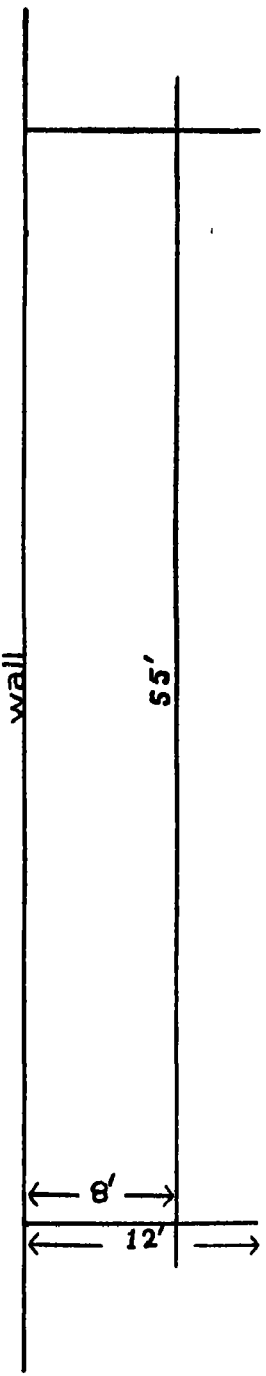


Figure 17. Floor Markings for Soccer Passing and Receiving Test

ball must be touched at least twice between each recovery and the next pass.

Practices: Each player is given one practice trial before the test begins.

Trials: The test consists of two trials from each end, the first two from A with the wall to the left, and the next two from B with the wall to the right.

Scoring: The score shall be the sum of the number of successful passes and recoveries made in the four trials. A successful pass and recovery shall be one in which the ball is kicked against the wall from outside the restraining line and first touched again with the foot outside the restraining line after the rebound.

Reliability: r = .56 by the odd-even method, when corrected by the Spearman-Brown formula, .72. The subjects were the same as for Test 1.

Validity: r = .50 with the subjective criterion, .68 with the total test criterion described for Test 1.

T-scores: See page 83.

Reference: Same as for Test 1.

Comments: This is a test involving difficult skills. The reliability and validity would both probably be higher for more experienced players. Further experimentation needs to be done on distances, length of time, number of trials, and perhaps even on amount of practice permitted. It is an excellent practice device.

3. JUDGMENT IN PASSING

Equipment: 1. One or more soccer balls, fully inflated; one stop watch; one bench six feet in length; one regulation soccer goal.

2. Markings

(a) Beginning from a point on the goal line and 4 feet inside the goal post, draw a line perpendicular to the goal line and

15 yards long. This is the restraining line.

(b) Extend the 12-yard line (regulation soccer marking) to a length equivalent to the length of the distance between the goal posts, and parallel to the goal line. This is the starting line.

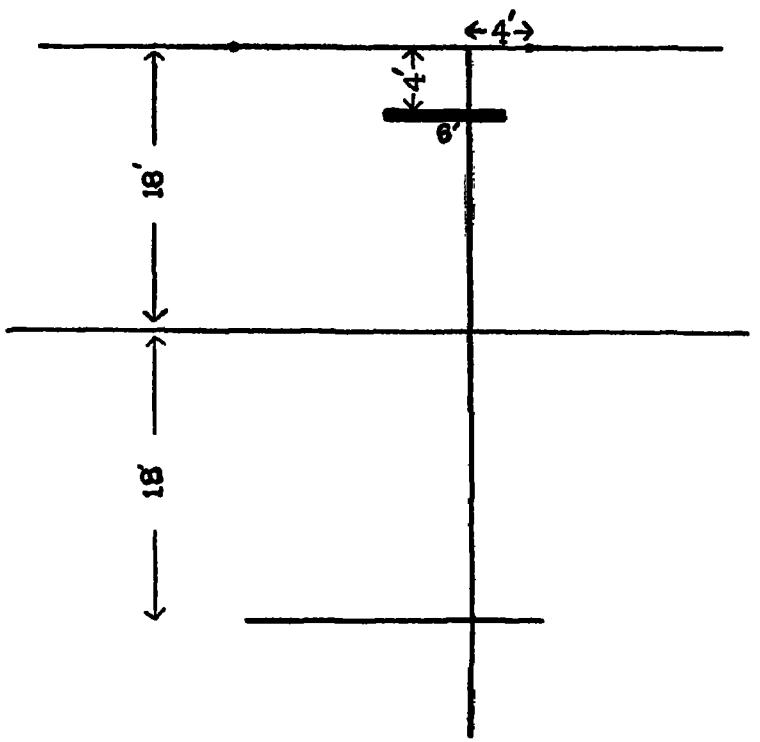


Figure 18. Field Markings for Soccer Judgment in Passing Test

This illustrates the restraining line and position of the bench for a kick from the right. The bench would be moved to the left and a restraining line marked on the left for the kick from the left.

(c) The 6-yard line, on all regulation fields, is also used in the test and is referred to as the 6-yard line.

(d) The bench is placed 4 feet from the goal, parallel to it, and in such a position

that one end is exactly at the center of the goal and the other extends toward the right side as you face the goal from the starting line.

(e) For the second part of the test, the above directions shall be reversed, and the only additional line needed is a restraining line, four feet inside the left hand goal post and perpendicular to the goal line, 15 yards long. (See Figure 18 for test from the right hand side.)

Test: Place the ball on the 12-yard line, outside the restraining line. On the signal, Ready, Go! dribble the ball forward, keeping to the right of the restraining line. After you have passed the 6-yard line, kick for goal when you think you are in the most advantageous position. You may use any kick you like, but the ball must be kicked from outside the restraining line and it must enter the goal between the left end of the bench and the left goal post. Each trial must be completed in four seconds, from the word "Go" until the ball is kicked. If you take longer or violate any of the other rules, the trial will have to be repeated. You are given five trials from the right side, then five from the left.

Instructions to test administrator: For speed in administering the test, three or more balls should be made available. Have each girl recover her own ball, carry it to the side of the field, and roll it to the person nearest the head of the line who does not have a ball. The entire test can be administered to as many as forty girls in forty minutes. Only five or six girls should be assigned to this test at one time.

Scoring: One point is given for each goal scored in a legal trial. Five legal trials are given as described above, and then the bench is set up on the other side of the goal and the entire test is repeated on that side. The score is the total for the ten trials.

Reliability: r = .69 when computed by the odd-even method. When the Spearman-Brown formula was applied, the reliability coefficient was raised to .82. The subjects were the same as for Test 1.

Validity: r = .34 with the subjective criterion; .65 with the total test criterion as described for Test 1.

T-scores: See page 83.

Reference: Same as Test 1.

Comments: This test is easily and quickly administered. The low validity figure is perhaps explained by the fact that judgment in placement and timing of passes is difficult to rate in an actual game situation.

4. Combination of Skills Test

Equipment: 1. Soccer balls, fully inflated; 2 goal posts or standards; one stop watch; playing surface 55 yards in length; and a wall space 12 feet long and 30 inches high. (Three locker room benches can be placed on their sides and stacked to present a smooth surface upon which the correct area can be painted with show card paint. They should be braced and stacked in place, so that there are no projections and so as to permit very little vibration.)

2. Field markings

(a) The soccer goal is 18 feet in width. Opposite the goal, mark a line 6 feet in length, parallel to the goal line and 55 yards from it. The center of this line should be directly opposite the center of the line between the goal posts. This is the starting line.

(b) Mark another line, to be used as a restraining line, extending from the left end of the starting line (as you face the goal) 25 yards toward the goal line.

(c) The "wall" is placed 12 feet to the left of this restraining line, so that the center of the marked target is directly opposite the end of the 25-yard line.

(d) A 6-yard line, parallel to the end line, is also used.

(e) The description above is for testing with the right foot. When testing with the left foot, the markings differ as shown in Figure 19. If the wall is stationary, new markings with the goal line as the starting point and the starting point as the goal line should be made.

Test: The ball is placed on the starting line. On the signal, Ready, Go! the ball is dribbled and kept to the right of the restraining line until you are at a point where you can kick the ball with your right foot (as though you were sending a pass to another person) and hit the wall. After the pass, run ahead to recover the ball as it rebounds and continue dribbling toward the goal until you are close enough to kick for goal. You must kick for goal before you cross the 6-yard line. Go as fast as you can without losing control of the ball. You are given four trials, with the wall to your left making the pass with the right foot; then there are four more trials, with the wall on your right, passing with the left foot.

Note to test administrators: Allow two practice trials with each foot. Record actual trials in half seconds. Trials which include errors must be repeated.

Scoring: Your score will be the total amount of time it takes you to complete the eight trials, each one timed separately.

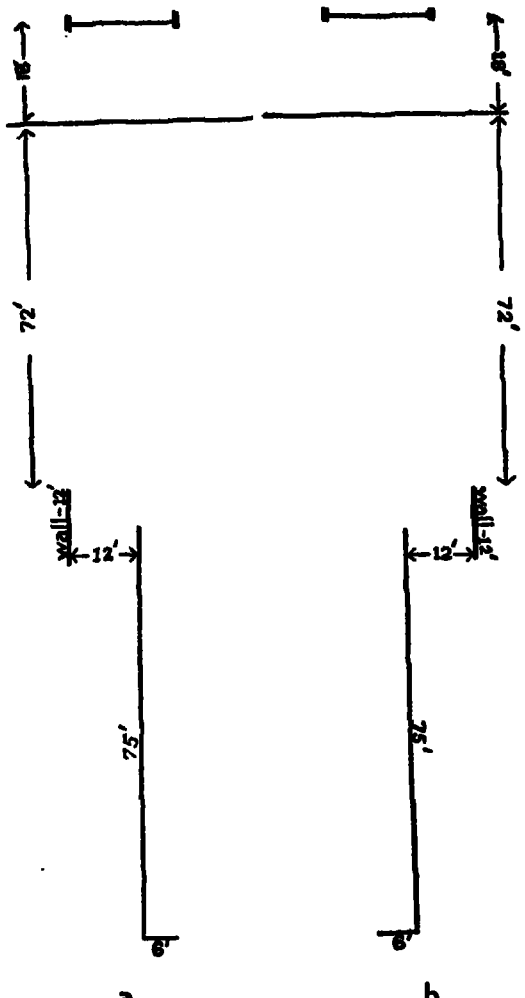


Figure 19. Field Markings for Soccer Combination of Skill Tests

a — markings for kick with the right foot
b — markings for kick with the left foot
||—|| — goal posts

Reliability: r = .93 computed by the odd-even method, when corrected by the Spearman-Brown formula, .96. The subjects were 124 fifth and sixth grade children from public

schools in Des Moines, Iowa, and in Webster Groves, Missouri.

Validity: For the 92 cases from Webster Groves, r = .92 with a subjective rating criterion. For the 32 cases from Des Moines, r = .53 with subjective ratings.

Scores: The range for fifth grade, 95.5 to 226.5 seconds; median 139.2.

The range for sixth grade, 90.0 to 191.0 seconds; median 130.3.

The range for girls only, 102.5 to 226.5 seconds; median, 143.4.

The range for boys only, 90.0 to 198.0 seconds; median 126.7.

T-scores: See page 83.

Reference: Bontz, Jean: An Experiment in the Construction of a Test for Measuring Ability in Some of the Fundamental Skills Used by Fifth and Sixth Grade Children in Soccer, Unpublished M.A. thesis, University of Iowa, 1942.

Comments: This test has proven interesting to players. If a stationary wall, such as a school building, is adjacent to the playing field, it can be practiced outside of class time. Such an arrangement leaves the playing field free. It has been used with high school and college ages, but data are not available on the older groups.

BATTERY OF SOCCER TESTS

The one best test appears to be volleying. A good two-item battery of tests is the combination of passing and receiving with volleying, which Schaufele found to give a coefficient of .85 with the total test criterion, or .63 with subjective ratings. The intercorrelation of these two tests is .46. The best weighting of the scores is:

2. passing and receiving + 1. volleying

An equally good two-item battery is the combination of judgment in passing with volleying, which give a coefficient of .85 with the total test criterion (subjective estimate figures not given). The intercorrelation of these two tests is .41. The recommended weighting of scores is:

2. judgment in passing + 1. volleying

The combination of the three tests gives a coefficient of .89, with the total test scores criterion. The correct weighting is:

2. passing and receiving + 1. volleying + . judgment in passing

TABLE V

T-SCALES FOR SOCCER TESTS

T-Scores	*Volleying*[a] *(253)*[c]	*Passing and Receiving*[b] *(84)*[c]	*Judgment in Passing*[b] *(84)*[c]	*T-Scores*	*T-Scores*	*Volleying*[a] *(253)*[c]	*Passing and Receiving*[b] *(84)*[c]	*Judgment in Passing*[b] *(84)*[c]	*T-Scores*
79	41			79	49	11			49
77	25			77	46	10	6	3	46
74	24			74	45				45
72	23			72	43	9			43
71	22		7	71	41		5		41
70	21	10		70	40	8		2	40
69	20			69	37	7			37
67	19		6	67	36		4		36
66	18			66	33	6			33
63	17	9		63	31	5		1	31
62			5	62	30				30
61	16			61					
					29		3		29
59	15			59	28	4			28
57	14	8		57	25	3	2		25
55	13			55	21	2			21
53	12			53					
51		7	4	51					

[a] Scale constructed on data collected on ninth and tenth grade girls, Fairview High School, Fairview Village, Ohio; and ninth grade girls in Blue Island Community High School, Blue Island, Illinois.
[b] Scale constructed on data collected on ninth and tenth grade girls, Fairview High School, Fairview Village, Ohio.
[c] Indicates the number of subjects included in the scale.

SOFTBALL

1. REPEATED THROWS

Equipment: 1. A number of new balls; a flat, unobstructed wall surface about 15 feet or more high and 8 feet wide; a stop watch.

2. Markings
Draw a line on the wall 7½ feet from the floor. Draw a restraining line on the floor 15 feet from the wall, and parallel to it.

Test: The player being tested stands any place behind the restraining line and facing the wall. On the signal, Ready, Go! the player throws the ball against the wall so that it hits above the 7½-foot line, catches it, and repeats this as many times as he can in 30 seconds. One ball is used throughout the test; if it gets out of control, it must be recovered by the player being tested. (The loss of time is considered sufficient penalty.) Foot faults (stepping on or over the line) are watched by the scorer and the player is told to move back. Any throws made while the player is on or over the line do not count. A rest of two minutes is recommended between trials; this can be easily administered if three or four girls take throws at the same target. Six trials are given.

Scoring: One point is counted each time the ball hits on or above the 7½-foot line, provided the throw was made when the player was behind the restraining line. The score for the entire test is the total of six trials of 30 seconds each.

Reliability: r = .89 by the odd-even method, .94 corrected by the Spearman-Brown formula; subjects were 66 college women. Underkofler found slightly lower coefficients on a similar test for junior high school girls. The 14-inch ball was used, the line on the wall was 10 feet high

and the restraining line was 10 feet from the wall. In this case the r was .73 for first and second trials, stepped up to .84 because the sum of the two trials was used as the score.

Validity: In the junior high school study cited above, the r with a subjective rating criterion was .64. In another study this test was given to 173 college women, in various institutions located in the central states. The coefficient was .51 with a subjective rating criterion. The comparatively low validity is doubtless partially explained by the fact that the same persons did not make all of the ratings. (See reference 1 below.)

T-scores: See page 88.

References: 1. Thomas, Jesselene: "Skill Tests," *Official Softball-Volley Ball Guide,* 1943, pp. 21-27. (Report of a project of the Research Committee of the Central Association of Physical Education for College Women, Aileen Carpenter, Chairman.) (Quoted by permission of A. S. Barnes & Co.)

2. Underkofler, Audrey: *A Study of Skill Tests for Evaluating the Ability of Junior High School Girls in Softball.* Unpublished M.A. thesis, University of Iowa, 1942.

Comments: This test is highly reliable but does not differentiate clearly between students in the middle ability group. The better players will be able to throw the ball with sufficient force so that it rebounds to them without bouncing. To a certain extent it measures accuracy as well as power. It correlates highly with the distance throw (intercorrelation .81) so there is no need to give both tests. It is easy to administer and can be given indoors on a rainy day. If you have sufficient wall space for

ten targets, the entire test can be administered to a class of forty girls in fifteen minutes.

2. DISTANCE THROW

Equipment: A number of regulation softballs, and a field. (See Figure 5 for field markings.)

Test: The player stands behind the line and throws the ball as far as possible with an overhand or sidearm motion. The player is limited to one step, which must be taken behind and not over the line. Three throws constitute one trial, and only the best throw of the three is measured and recorded. Three trials are permitted (9 throws in all).

Scoring: The throw is measured as the distance in feet from the starting line to the spot where the ball first touches the ground. The best of the three recorded throws is used as the player's score.

Reliability: r = .95, computed on successive trials. The subjects were 118 girls in the seventh and eighth grades in the Intermediate School, Riverside, Illinois.

Validity: r = .63 with ratings, subjects were college girls (same group as for Test 1).
r = .81 with ratings, subjects were 118 seventh and eighth grade girls.

T-scores: See page 88.

References: Same as for Test 1.

Comments: Ability to throw the ball long distances is important in softball, and since there is a relationship between the distance that the ball can be thrown and the ability to throw the ball with speed, it seems all the more important that this skill be measured. This appears to be the best single test yet devised for softball playing ability.

3. BATTING

(*Note:* The study of the test described below was made using a fourteen-inch softball, rather than the regulation twelve-inch ball. The pitching distance was 30 feet rather than the official 35 feet for girls. The 45-foot diamond was used, rather than the 60-foot. These points should be kept in mind in interpreting results.)

Equipment: 1. Several fourteen-inch balls, and regulation softball bats.

2. Markings

A 45-foot diamond, with infield marked; 30-foot pitcher's distance.

Test: The contestant stands in the batter's box and is given 10 trials to hit legal pitches which are called strikes by an umpire. "Balls" are disregarded. The effect of the personal element is held somewhat constant by using the same pitcher and umpire throughout the administration of the test.

Score: Five points for each outfield hit, 3 points for an infield hit, 1 point for a foul, and 0 points for a ball struck at and missed or for a called strike. The score is the total number of points earned in the ten trials.

Reliability: $r = .65$ by the odd-even method, raised to .79 with the Spearman-Brown correction. The subjects were 118 girls in seventh and eighth grades.

Validity: The validity coefficient of this test was found to be .72 for the group described above. The criterion was a subjective rating, made by two students in each class, plus two times the rating score given by the instructor.

T-scores: See page 88.

Reference: Underkofler. (Same as for Test 1.)

Comments: Since batting comprises such a large share of the offensive play in softball, no battery of tests seems complete without it. This is the

first test that presents any statistical data. It is recognized that even though the same pitcher is used that the conditions are not the same for all batters as that pitcher will vary in his deliveries. Before much confidence can be placed in the test, further study is desirable with the test administered with official balls, pitching distance, and diamond. The use of a nationally rated softball umpire would also be helpful.

BATTERY OF SOFTBALL TESTS

The best single test appears to be the distance throw. The intercorrelation between Underkofler's batting test and her distance throw is relatively high (.63). When the two are combined

TABLE VI

T-SCALES FOR SOFTBALL TESTS

T-Score	*Repeated Throws (118)* [a]	*Repeated Throws (66)* [b]	*Distance Throw* [c]	*Batting (118)* [d]	*T-Score*
76	37		131	42	76
75			129		75
74			127		74
73			126		73
72		116-117	124	40	72
71			123		71
70	36		121		70
69			120		69
68			118	38	68
67	35		117		67
66		114-115	115	36	66
65			114	35	65
64	34		112		64
63			110		63
62	33	112-113	109	34	62
61	32	110-111	107		61
60		108-109	106	32	60
59			104	30	59
58		106-107	103	29	58
57	31		101	28	57
56		104-105	100	27	56

TABLE VI (Continued)

T-Scales for Softball Tests

T-Score	*Repeated Throws (118)* [a] [d]	*Repeated Throws (66)* [b] [d]	*Distance Throw* [c]	*Batting (118)* [a] [d]	*T-Score*
55	30	100-103	98	26	55
54		98-99	96	24	54
53	29	96-97	95	23	53
52			93	22	52
51		94-95	92		51
50	28	92-93	90	21	50
49		90-91	87	20	49
48		88-89	85	19	48
47	27	86-87	82		47
46		82-85	80	18	46
45	26		78	17	45
44	25	80-81	75	16	44
43		76-79	73		43
42	23	74-75	70	15	42
41	22	72-73	68	14	41
40	21	70-71	65		40
39	19		63	13	39
38	18		60	12	38
37	16	68-69	58	11	37
36	14	66-67	55	9	36
35	13	64-65	53	8	35
34	10		50		34
33		62-63	48		33
32	9		46	6	32
31	7	60-61	43		31
30		58-59	41		30
29			40		29
28	4			1	28
27					27
26		56-57			26
25					25
24	2				24

[a] Scale constructed on data collected on seventh and eighth grade girls in Riverside School, Riverside, Illinois. For form of the tests see discussion on reliability of Test 1.) Score is the sum of two trials.
[b] Scale constructed on data obtained on college women at University of Iowa, and University of Minnesota. Score is sum of six trials.
[c] Scale constructed on data obtained on college women. (Quoted by permission of A. S. Barnes & Co.)
[d] Indicates number of cases represented in the scale.

the correlation is not high enough to warrant giving the second test. Since the distance throw and repeated throws have been shown to measure quite similar abilities, no combination is possible there.

Various base running tests have been tried but it seems to be a relatively unimportant skill to include in a limited battery. It is known that higher validities are obtained when the test includes running from home plate to second base than in running to first base only or in running all four bases. Perhaps the best battery to recommend at present is a combination of ratings of batting skill (see Chapter 7) with either the distance throw or repeated throws.

SPEEDBALL

1. LIFT TO OTHERS

Equipment: 1. Soccer balls, net (volley ball, badminton, or tennis), and standards to hold the net at 2½ foot height.

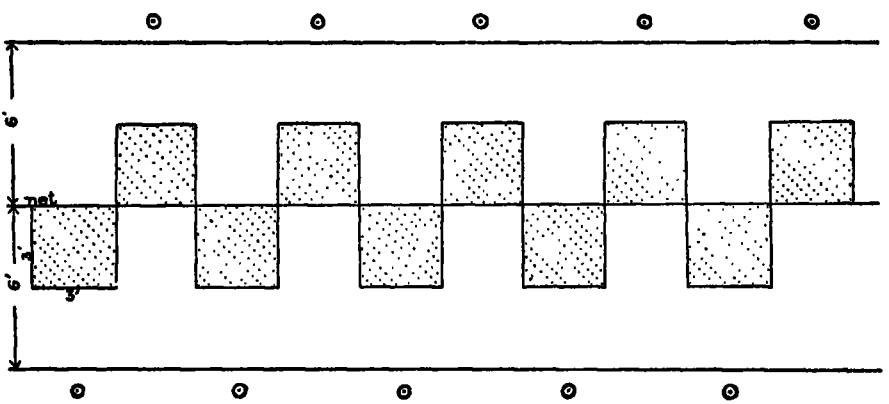


Figure 20. Field Markings for Speedball Lift to Others Test

⊙ players in position ready to take the test

2. Markings

One line on each side of the net, parallel to the net and 6 feet from it, and extending the length of the net. Three-foot squares must be drawn at 3-foot intervals all along both sides of the net. (See Figure 20.)

Test: Each player stands behind the 6-foot line directly behind a 3-foot square, with a soccer ball placed on the line. The player then lifts the ball with either foot, attempting to send the ball across the net and into the square diagonally opposite him and to his right. The player on the opposite side recovers the ball and lifts the ball back to the 3-foot square diagonally to his right and across the net. If the ball touches the net and goes over and lands within the designated area, the score is counted. Five trials to the right and five to the left are given. Partners on opposite sides of the net keep each other's scores and report to student scorers.

Scoring: The total number of correct lifts out of ten trials is the score, each accurate and successful lift counting one point.

Reliability: The reliability, by correlating odd and even trials, was .87. When the Spearman-Brown Prophecy Formula was applied, to step the test up to actual length, the reliability was .93. The subjects were 72 high school girls, Parsons, Kansas. The tests were given at the end of the season.

Validity: The validity of this test was found to be .88 for the group described above. The criterion was subjective ratings, by three teachers of speedball playing ability.

T-scores: See page 99.

Reference: Buchanan, Ruth E.: *A Study of Achievement Tests in Speedball for High School Girls.* Unpublished M.A. thesis, University of Iowa, 1942.

Comments: The lift of a stationary ball in speedball is used frequently to begin the game and to start play after each violation. Thus, it would seem to be an important enough skill to be

included in most test batteries. Unless players have practiced this skill a great deal, and are practically at the peak of their performance ability, one could not expect such high reliability.

2. THROWING AND CATCHING WHILE STANDING

Equipment: 1. Soccer balls, wall space, stop watch.

2. Markings

Line drawn 6 feet from the wall and parallel with it, extending the length of an uninterrupted wall space.

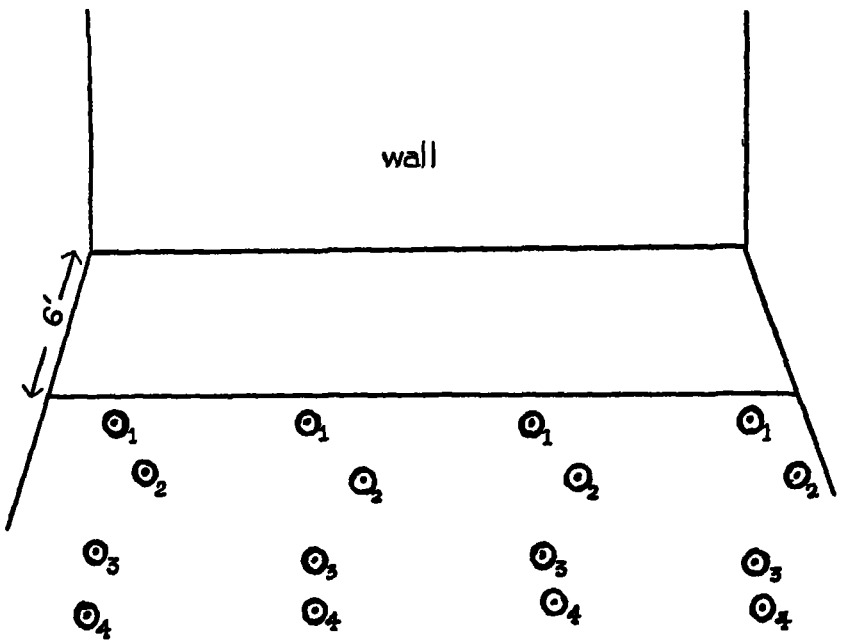


Fig. 21. Floor Markings for Throwing and Catching While Standing Test

$\odot_1$ players in position ready to take test
$\odot_2$ players in position ready to score test
$\odot_3$, $\odot_4$ players in position waiting their turns at the test

Test: Each player stands behind the 6-foot line with a soccer ball in his hands. The scorer for each player is stationed to the rear and slightly to one side. At a signal, the player begins by throwing the ball to the wall, continues by catch-

ing and repeating. Each trial is 30 seconds in length. Each player is given 5 trials, but these trials should not be in succession.

Scoring: Each throw made from behind the line counts one point. The total for each trial is recorded and the score for the test is the average for the five trials.

Reliability: By the odd-even method the coefficient was .85. When corrected by the Spearman-Brown Prophecy Formula, the reliability was .92. The subjects were the same as for Test 1.

Validity: The coefficient was .79; the criterion was a rating by experts; the subjects were the same as for Test 1.

Reference: Same as for Test 1.

T-scores: See page 99.

Comments: This test can be administered in a very short time. If the players are in lines of three or four, and rotate in taking turns, the element of fatigue is prevented from affecting the scores. The test author suggests that because of the high reliability, the number of trials might be reduced to as few as two. This is very similar to the wall pass test in basketball, and substantiates the evidence there that very few trials are required. The high reliability is typical of tests involving time. Note the same tendency in softball "repeated throws" and in volley ball "repeated volleys."

3. KICK-UPS

Equipment: 1. Soccer balls.

2. Markings

Three feet out from the side line, mark a line parallel to the sideline and 2 feet in length. Complete a square, with 2-foot sides, as shown in the diagram below. Locate a point which is on the imaginary extension of

a diagonal line across the square and 4 feet from the nearest corner. Mark this spot and use it as a starting line. Arrange as many similar targets as needed for administering the test efficiently.

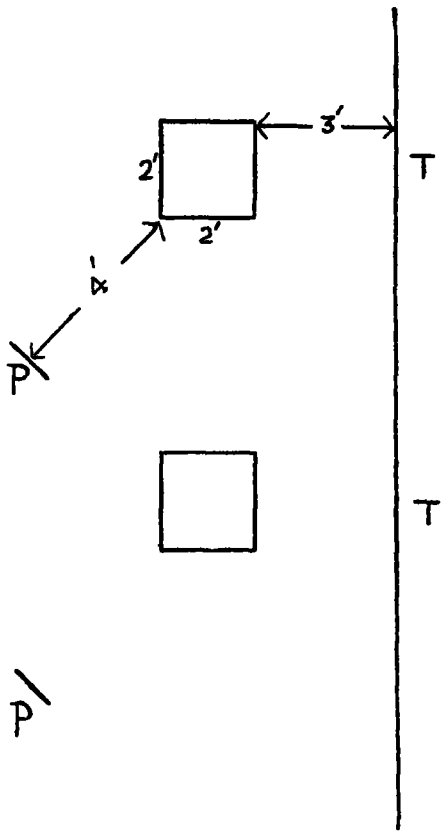


Figure 22. Field Markings for Speedball Kick Up Tests

T — thrower, ready to put ball in play
P — player in starting position ready to take test

Test: The player being tested stands behind the starting line. The thrower takes a position behind the sideline and drops the ball in the 2-foot square. As soon as the thrower releases the ball, the player being tested runs forward to execute a kick-up. Five practices are given followed by ten trials.

Scoring: The total number of successfully executed and caught kick-ups of the ten trials is the score. If the ball is not thrown so that it lands in the 2-foot square, the trial is repeated.

Reliability: By the odd-even method, the coefficient was .87; corrected by the Spearman-Brown formula .93. The subjects were the same as described in Test 1.

Validity: r = .85.

T-scores: See page 99.

Reference: Same as Test 1.

Comments: This test has proven quite challenging to players and it is a skill that is practiced a great deal by all players, regardless of their position. It includes the skills of judging the speed and height of the ball, the ability to execute the kick-up, and to catch and hold the ball.

4. DRIBBLING AND PASSING

Equipment: 1. Soccer balls and five objects (Indian clubs, jump standards).

2. Markings

Mark a starting line, 60 yards from one of the end lines of the field and parallel to it. Place one object 3 yards to the right of the right hand goal post (as you face the goal from the center of the field), and 10 yards from the end line.

Place the other objects at 10 yard intervals in a line parallel to the sideline, so that one object is 10 yards from the starting line.

If you wish to give the test to as many as four players at one time, use an arrangement similar to that on the diagram, marking the goal space with 12-inch lines, perpendicular to the end line. The two set-ups on the right, as you face the goal line, will permit adminis-

tration of the test with the pass to left, while the two on the left are for the pass to the right.

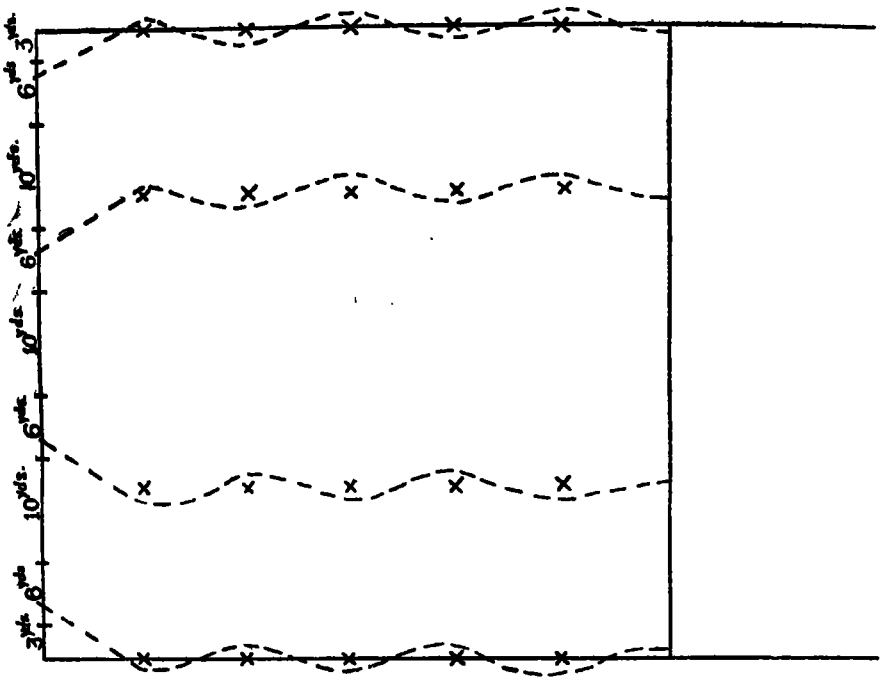


Figure 23. Field Markings and Plan for Administering Speedball Dribbling and Passing Tests

x — obstacles
Field is marked for two courses kicking from the right and two from the left

Test: Ball is placed on the starting line, in line with the objects. On signal, the player starts dribbling forward, going to the right of object 1, to the left of #2, and so on, and to the right of object #5, from which place he immediately passes the ball to the left, attempting to send it between the goal posts (or indications of goal posts). There are five trials; then five more trials from the other side, where it is necessary to pass to the right to send the ball through the goal area. To do this, the player must go to the right of the

first object. This makes a total of ten trials.

Scoring: The number of seconds to dribble 50 yards ten times, minus ten times the total number of accurate passes.

Reliability: Odd-even r = .96, corrected by Spearman-Brown formula = .98. The subjects were the same as for Test 1.

Validity: r = .69, with a criterion of ratings described in Test 1.

Reference: Same as for Test 1.

Comments: This test would be improved with a little more refinement. Also, a line 10 yards from the end line and parallel to it, passing under the fifth object, would increase the accuracy in timing. The addition of the requirement that the ball be passed on or before reaching the 6-yard line would prevent players from going too near before making the pass. This test will take longer to administer than any of the others. It is possible that further study will show that the number of trials from each side could well be reduced to three, perhaps even two, and still be highly reliable.

PASSING

Equipment: Same as for Test 4 (dribbling and passing).

Test: Same as Test 4.

Scoring: The number of accurate passes out of ten trials.

Reliability: r = .84 by the odd-even method, .91 when corrected by the Spearman-Brown formula. Same subjects were used as for Test 1.

Validity: r = .86 with the criterion of ratings.

T-scores: See page 99.

Reference: Same as for Test 1.

Comments: This method of scoring results in higher validity. Further study should be done to determine the effect of shortening the distance required for dribbling before the pass is made.

6. DRIBBLING

Equipment: Same as for Test 4 (dribbling and passing).

Test: Same as Test 4.

Scoring: The total number of seconds to dribble 50 yards 10 times.

Reliability: r = .96 for the odd-even method, .98 corrected by the Spearman-Brown formula. The subjects were the same as for Test 1.

Validity: r = .57.

Reference: Same as for Test 1.

Comments: Dribbling ability is apparently not closely related to general playing ability. This test is included here to illustrate how a test can be highly reliable and still not have satisfactory validity; also, how the elements in Test 4 (dribbling and passing) were studied until one feels safe in concluding that passing the ball while moving is the important element of the test.

BATTERY OF SPEEDBALL TESTS

If time does not permit giving all the tests, the best two-item battery is throwing and catching while standing and the passing test. Buchanan found that this combination gave a high correlation (R = .93) with the criterion (general playing ability, judged by three teachers). This battery can be administered to a class of forty girls in forty minutes, provided that a set-up is arranged for the passing test similar to that shown in Figure 22. The proper weighting for this two-item battery is

1. throwing and catching + 3. passing

That the two tests do not measure exactly the same skill is shown in the relatively low intercorrelation of .57.

TABLE VII

T-Scales for Speedball Tests [a]

LIFT TO OTHERS (262) [b] 10 trials		THROWING AND CATCHING STANDING (72) [b] Mean of 5 trials		THROWING AND CATCHING STANDING (159) [b] Mean of 2 trials	
Raw Score	*T-score*	*Raw score*	*T-score*	*Raw score*	*T-score*
10	71	19.8-20.2	75	27.8-29.7	72
9	63	19.3-19.7	72	25.8-27.7	65
8	57	18.8-19.2	69	23.8-25.7	57
7	53	18.3-18.7	67	21.8-23.7	52
6	47	17.8-18.2	65	19.8-21.7	48
5	43	17.3-17.7	62	17.8-19.7	45
4	39	16.8-17.2	60	15.8-17.7	41
3	34	16.3-16.7	58	13.8-15.7	37
2	30	15.8-16.2	56	11.8-13.7	33
1	25	15.3-15.7	54	9.8-11.7	26
		14.8-15.2	51		
		14.3-14.7	46		
		13.8-14.2	43		
		13.3-13.7	42		
		12.8-13.2	41		
		12.3-12.7	40		
		11.8-11.2	39		
		11.3-11.7	36		

KICK-UPS (262) [b] 10 trials		PASSING (190) [b] 5 trials		PASSING (72) [b] 10 trials	
Raw Score	*T-score*	*Raw Score*	*T-score*	*Raw Score*	*T-score*
10	72	5	68	10	71
9	66	4	51	9	66
8	59	3	43	8	62
7	54	2	35	7	59
6	50	1	28	6	57
5	47			5	51
4	43			4	46
3	38			3	40
2	34			2	35
1	29			1	25
0	21				

[a] The scales were constructed from data obtained in the Buchanan study on high school girls in Parsons, Kansas.

[b] The number of subjects included in each scale is indicated at the top.

The scale on throwing and catching standing (mean of 2 trials) illustrates especially well the effect of making the step-intervals too large, resulting in large gaps in the T-scale. The range in all the tests is small which accounts for the gaps and irregularities in each of the scales.

TENNIS

1. WALLBOARD TEST

Equipment: 1. Backboard or wall, approximately ten feet in height and allowing fifteen to twenty feet in width per person taking the test at one time; stop watch; two balls and a racquet per player. Balls should be in good condition and racquet should be tightly strung. Box for extra balls, about 12 inches long, 9 inches wide, and 3 inches deep, placed on the floor where the restraining line (described below) joins the side (at the left for right handed players and right for left handed players). A racquet may be substituted for the box; the racquet is placed on the floor in the same position as that described for the box, and the balls are laid on the face of the racquet.

2. Markings

A line, 3 inches in width, should be drawn on the wall, to represent the net, so that the top is 3 feet from the floor. A restraining line, 5 feet from the base of the wall should be drawn on the floor, parallel to the wall. (See comments below for change in position of the restraining line.)

Test: On the word *Go* of the signal, Ready, Go! drop the ball and let it hit the floor once and then start rallying it against the wall. Continue rallying until the signal to stop. The ball may bounce any number of times or it may be volleyed. At the start of the test and whenever a new ball is put in play, it must be allowed to bounce before being hit. Any stroke may be used but all strokes should be played from behind the restraining line. You may cross the line to retrieve the ball but hits made from this position are not scored.

If the ball gets out of control, you may take another ball from the box. (See p. 43 for a discussion of the demonstration for this test.)

Scoring: Each time a ball strikes the wall on or above the net line, having been hit from behind the restraining line, one point is scored. Three trials are given, and the score is the sum. The length of each trial is 30 seconds.

Reliability: Test appears highly reliable.

Validity: Using the criteria of rankings by three experts and of round robin tournament rankings, the correlations range from .90 to .92, according to Dyer.

Reference: Dyer, Joanna Thayer: "Revision of the Backboard Test of Tennis Ability," *Research Quarterly*, 9, March, 1938, p. 25. (Quoted by permission of the *Research Quarterly*.)

Comments: This test is satisfactory for a classification test. It should not be used to measure form in tennis. It can be administered easily and is completely objective.

We recommend moving the restraining line to a distance of 20 or 25 feet from the wall (distance depends somewhat on the wall surface and the type of rebound it gives). See comments in Chapter 2, p. 15.

2. TESTS FOR FORM

See Chapter 7 for suggestions on rating tennis form. No objective tests are available.

VOLLEY BALL

1. REPEATED VOLLEYS

Equipment: 1. Well inflated balls, unobstructed wall space 10 feet long and 15 feet high (preferably several such areas), and a stop watch.

2. Markings

A line 10 feet long marked on the wall at net height, 7½ feet from the floor.

A line on the floor, opposite the wall marking, 10 feet long and 3 feet from the wall.

Test: The player being tested shall stand behind the 3-foot line, and shall toss the ball to the wall with an underhand movement. When it returns, she shall volley it repeatedly against the wall above the net line for 15 seconds. The ball may be set up as many times as desired or necessary; it may be caught and re-started with a toss as at the beginning. If the ball gets out of control, it must be recovered by the subject and brought back to the 3-foot line to be started over again as at the beginning.

This procedure shall be repeated until ten trials have been given, each 15 seconds in length.

Scoring: The score for one trial shall be the number of times the ball is clearly batted (not tossed) from behind the 3-foot line on the floor to the wall above or on the net line. The score for the test shall be the sum of the five best trials out of ten.

Reliability: r = .78, correlating by the odd-even method. It was not corrected by the Spearman-Brown formula since only five of the trials are used in the final score. Reliability computed on the five best trials on successive days should yield an equally high, perhaps higher, coefficient. The subjects were 47 senior high school girls, University High School, Iowa City, Iowa.

r = .96 computed by the odd-even method; subjects were 75 college women, University of Minnesota. Ten trials; all used.

Validity: r=.72 when correlated with a criterion of subjective ratings, made by four experienced teachers of volley ball.

T-scores: See page 105.

Reference: French, Esther and Cooper, Bernice: Achievement Tests in Volley Ball for High School Girls, *Research Quarterly,* 8, May, 1937, p. 150. (Quoted by permission of the *Research Quarterly.*)

Comments: This test measures the player's ability to control the ball, and his judgment on position and playing the ball. It can be used for classification purposes early in the term and then again later to measure achievement. It can be administered economically with a large number of players taking the test at the same time, if wall space and balls permit.

2. SERVE

Equipment: 1. Regulation court and tightly strung net, well inflated balls.

2. Markings

(a) A line across the court 5 feet inside and parallel to the end line.

(b) A line across the court parallel to the net and 12½ feet from the center line which is directly under the net.

(c) Two lines each 5 feet inside the court and parallel to the side lines, extending from the line under the net to the 5-foot line described in (a).

(d) The score values of each area should be marked on the floor as indicated in the diagram (Figure 24).

Test: The player being tested stands in the proper serving area on the court opposite the target and is given ten trials to serve the ball into the target in the court across the net. Any legal serve is permitted. Foot faults shall count as

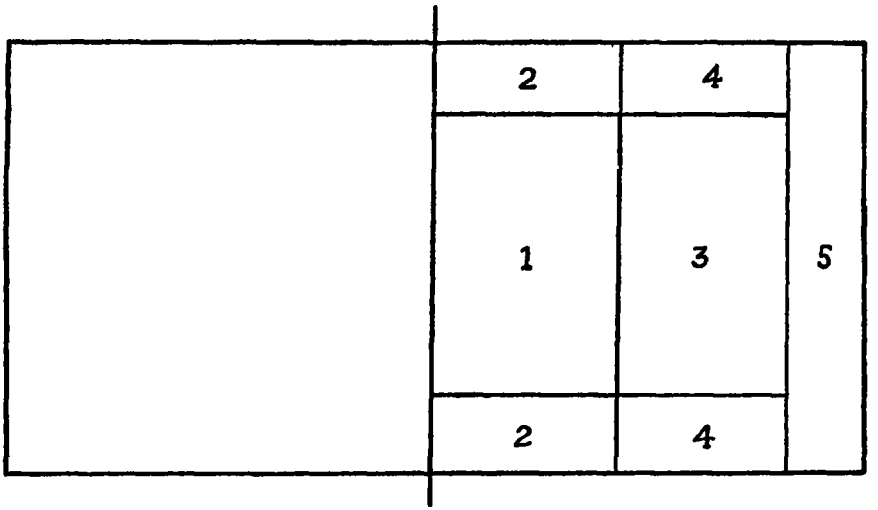


Figure 24. Floor Markings for Volleyball Serve Test

1 – 5-score for respective areas

trials; "let" serves shall be reserved and do not count as trials. The scorer stands on a chair near one sideline about fifteeen feet from the net.

Scoring: The score values are indicated on the diagram. A ball landing on a line separating the two spaces scores the higher value. A ball landing on an outside boundary line scores the value of the area the line bounds. Trials in which foot faults occur score zero. Twenty trials should be allowed.

Reliability: r = .68 by the odd-even method, stepped up to .81 by the Spearman-Brown formula. The subjects were the same as for Test 1. (University High School, Iowa City.)

Validity: r = .63, with a criterion of ratings (same as Test 1).

T-scores: See page 105.

Reference: Same as for Test 1.

Comments: This test is time consuming, since it requires so many trials. When the target is painted on both sides of the court, two players can be tested at the same time on each court. The

use of ball chasers to keep the servers supplied with balls facilitates the testing. It is an excellent teaching device for practice of placement in serving.

BATTERY OF VOLLEY BALL TESTS

The best single test is the repeated volleys. The two tests together yielded a correlation coefficient with the criterion (rating of playing ability) of .81 in the study cited here. This combination can be administered at the same time, if floor space permits. The low intercorrelation of .39 between the two tests indicates that they measure different things. The formula for combining the two tests, to insure giving proper weight to each, is as follows:

1. repeated volleys + 2. serve

TABLE VIII

T-SCALES FOR VOLLEY BALL TESTS

T-Score	*Repeated Volleys (349)* [a]	*Repeated Volleys (120)* [b]	*Repeated Volleys (136)* [c]	*Serve (349)* [d]	*Serve (142)* [e]	*T-Score*
80	130-139			48		80
79						79
78	126-129			47	44	78
77			40			77
76		25		46		76
75	124-125			45		75
74	122-123			44		74
73	120-121		38	43	40	73
72	118-119	24		42		72
71	116-117	23	37		35	71
70		22	35	41	34	70
69	114-115	21	34	40		69
68	110-113		33	39	33	68
67	108-109		32	38		67
66	106-107	20		37	32	66
65	104-105	19	31	36	31	65
64	102-103	18	30	34	30	64
63	100-101	17	29	33		63
62	98-99		27	32	29	62
61	96-97		26	30	28	61
60	94-95	16	24	28	27	60
59	92-93		22	27	26	59
58		15	21	26	25	58
57	90-91		20	25	24	57
56	88-89	14	19	24	23	56

TABLE VIII (Continued)

T-Scales for Volley Ball Tests

T-Score	*Repeated Volleys (349)* [f] [a]	*Repeated Volleys (120)* [f] [b]	*Repeated Volleys (136)* [f] [c]	*Serve (349)* [f] [d]	*Serve (142)* [f] [e]	*T-Score*
55	86-87		18	23	22	55
54		13		22	21	54
53	84-85		17	21	20	53
52	82-83	12		20	19	52
51	78-81		16	19	18	51
50	76-77		15	17	14	50
49		11		16	13	49
48	74-75		14	15	12	48
47	72-73			14	11	47
46	70-71	10		13	10	46
45	68-69		13	12	9	45
44	64-67			10		44
43	62-63	9	12	9	8	43
42	58-61				7	42
41	56-57	8	11	8		41
40	52-55		10	7		40
39	48-51			6	6	39
38	44-47	7	9	5	5	38
37	42-43					37
36	40-41			4	4	36
35	38-39			3		35
34	36-37	6	8		3	34
33	34-35			2		33
32	32-33		7	1		32
31	30-31	5			2	31
30	28-29					30
29	24-27	4				29
28						28
27	22-23		6			27
26	20-21					26
25	18-19					25
24	16-17	3				24
23						23
22						22
21	14-15					21

[a] Scale constructed from data obtained on girls in Blue Island Community High School, Blue Island, Illinois, and in East Aurora High School, Aurora, Illinois. Score is the sum of the best five out of ten trials.
[b] Scale constructed on data obtained from girls in Muscatine High School, Muscatine, Iowa; score is the best single score out of three trials.
[c] Scale constructed on data obtained on University of Iowa students. Score is the best of three trials.
[d] Same group as in a; score is total of 10 trials.
[e] Same group as in c; score is total of 10 trials.
[f] Indicates the number of cases in the distribution.

BIBLIOGRAPHY

1. Bassett, Gladys, Glassow, Ruth, and Locke, Mabel: "Studies in Testing Volley Ball Skills," *Research Quarterly,* 8, December 1937, p. 60
2. Bontz, Jean: *An Experiment in the Construction of a Test for Measuring Ability in Some of the Fundamental Skills Used by Fifth and Sixth Grade Children in Soccer.* Unpublished M.A. thesis, University of Iowa, 1942
3. Brace, David K.: "Validity of Football Achievement Tests as Measures of Motor Learning and as a Partial Basis for the Selection of Players," *Research Quarterly,* 14, December 1943, p. 372
4. Brechler, Paul W.: *A Test to Determine Potential Ability in Football (Backs and Ends).* Unpublished M.A. thesis, University of Iowa, 1940
5. Buchanan, Ruth E.: *A Study of Achievement Tests in Speedball for High School Girls.* Unpublished M.A. thesis, University of Iowa, 1942
6. Colvin, Valerie, Glassow, Ruth B., and Schwartz, Marguerite: "Studies in Measuring Basketball Playing Ability of College Women," *Research Quarterly,* 10, October 1938, p. 60
7. Cormack, Herbert P.: *A Test to Determine Potential Ability in Football (Linemen).* Unpublished M.A. thesis, University of Iowa, 1940
8. Cozens, Frederick, Cubberley, Hazel, and Neilson, N. P.: *Achievement Scales in Physical Education Activities for Secondary School Girls and College Women.* A. S. Barnes & Company, New York, 1937
9. Cozens, Frederick, Trieb, Martin H., and Neilson, N. P.: *Physical Education Achievement Scales for Boys in Secondary Schools.* A. S. Barnes & Company, New York, 1936
10. Dyer, Joanna Thayer: "Revision of the Backboard Test of Tennis Ability," *Research Quarterly,* 9, March 1938, p. 25
11. Dyer, Joanna Thayer, Schurig, Jennie C., and Apgar, Sara L.: "A Basketball Motor Ability Test for College Women and Secondary School Girls," *Research Quarterly,* 10, October 1939, p. 128
12. Edgren, H. D.: "An Experiment in the Testing of Ability and Progress in Basketball," *Research Quarterly,* 3, March 1932, p. 159
13. French, Esther, and Cooper, Bernice: "Achievement Tests in Volley Ball for High School Girls," *Research Quarterly,* 8, May 1937, p. 150
14. Glassow, Ruth B., and Broer, Marion R.: *Measuring Achievement in Physical Education.* W. B. Saunders Company, Philadelphia, 1938
15. Hewitt, Jack E.: "Achievement Scale Scores for Wartime Swimming," *Research Quarterly,* 14, December 1943, p. 391
16. Howland, Amy R.: *National Achievement Standards for Girls,* National Recreation Association, New York, 1936
17. Hyde, Edith I.: "An Achievement Scale in Archery," *Research Quarterly,* 8, May 1937, p. 109
18. Johnson, E. L.: *Objective Tests in Basketball for Boys.* Unpublished M.A. thesis, University of Iowa, 1934
19. Jones, Edith: *A Study of Knowledge and Playing Ability in Basketball for High School Girls.* Unpublished M.A. thesis, University of Iowa, 1941
20. Mitchell, A. Viola: "A Scoring Table for College Women in the Fifty-yard Dash, the Running Broad Jump, and the Basketball Throw," *Research Quarterly,* 5, Supplement March 1934, p. 86
21. Neilson, N. P., and Cozens, Frederick: *Achievement Scales in Physical Education Activities for Boys and Girls in Elementary and Junior High School.* A. S. Barnes & Company, New York, 1934

22. Schaufele, Evelyn F.: *The Establishment of Objective Tests for Girls of the Ninth and Tenth Grades to Determine Soccer Ability.* Unpublished M.A. thesis, University of Iowa, 1940
23. Schmithals, Margaret, and French, Esther: "Achievement Tests in Field Hockey for College Women," *Research Quarterly,* 11, October 1940, p. 84
24. Scott, M. Gladys: "Achievement Examinations in Badminton," *Research Quarterly,* 12, May 1941, p. 242
25. ——: "Assessment of Motor Ability of College Women through Objective Tests," *Research Quarterly,* 10, October 1939, p. 63.
26. Thomas, Jesselene: "Skill Tests," *Official Softball-Volley Ball Guide.* A. S. Barnes & Company, New York, 1943, p. 21.
27. Underkofler, Audrey: *A Study of Skill Tests for Evaluating the Ability of Junior High School Girls in Softball.* Unpublished M.A. thesis, University of Iowa, 1942
28. Young, Genevieve, and Moser, Helen: "A Short Battery of Tests to Measure Playing Ability in Women's Basketball," *Research Quarterly,* 5, May 1934, p. 3

5·
Evaluation of Physical Fitness

THE NEED FOR FITNESS TESTS

As a direct result of the war and war-time emphases, teachers have been forced to consider the outcome of their programs more carefully than they have ever done before. The requirements for waging total war have placed more and more of a premium on physical fitness. Physical fitness rose high in discussion of physical education objectives. Objectives long held were consolidated under a single composite term, physical fitness or total fitness. Immediately teachers, pupils and laymen raised their respective questions, "Am I achieving fitness in students?" "How fit am I?" or "Am I improving?" and "Is physical education really contributing to fitness?" Each question was the result of varying degrees of interest, curiosity and doubt. Each question deserved an answer.

Both student and layman expected a specific answer. We talked about fitness as an entity, therefore, it seemed justifiable to expect a single answer. They have become conditioned to high-power testing procedures with a composite score. Our I.Q. is expressed in a single figure; our motor quotient corresponds; the state-wide every pupil test gives the pupil or school a percentile ranking; the army gives an intelligence score; the popular magazines give the reader a self-administering test with a score on his powers of observation, or wit, or what have you. In such cases, the qualities measured are often more intangible than that of fitness and the scores are presented as a valid measure of that quality even in cases not entirely justified.

A survey of the history of testing reveals the development of highly specific forms of measuring devices. Tests are designed to measure motor ability, innate capacity or educability, size, strength, or a specific sport skill. The ones claiming to measure physical fitness were of two types. The first type was based on the physiological

work principles and consisted principally of measures of cardiovascular or respiratory functioning. Most of them involved highly complex apparatus, required administration as an individual test although they showed such low reliability that they seemed better adapted to group than to individual interpretation. The second type purporting to measure fitness included strength tests in their various forms. In spite of the claims for strength tests as indicators of health, vigor and fitness, most teachers considered these claims as exaggerated and considered them as another specific, a strength measure.

So the answer to the question of how to measure fitness did not seem to be forthcoming. It was apparently another case of "water, water everywhere and not a drop to drink," i.e., tests of all kinds but none suitable for this particular need.

Then as thinking on fitness became more crystallized and situations arose which illustrated the need for different kinds of fitness, these specific tests seemed to be more pertinent. The first reaction to that realization was the general policy of accepting one or two tests and admitting that they were inadequate but all that was available. The result was as many different systems as there were teachers. Evidence of this situation is to be found in the manual published by the United States Office of Education[21] early in 1943. With reference to the girls' program the manual says that there is a lack of scientific evidence available on tests for girls, and that standards presented represent the best available and may serve as guides to teachers and as incentives to pupils. The measures then suggested are jump and reach, a potato race, soccer throw-in, and a free style swimming event.

Even with the acceptance on an empirical basis of the available tests the problem was not solved. To be useful any test must have some standard, some basis for comparing the individual or the class. The same procedure has been followed again and again. For example, the Minnesota State Civilian Defense Council[18] quickly set up suggested standards for both sexes, from age 12 up to 35 years and up, in order to get the state-wide Physical Fitness Club going. A similar purpose was behind the writing of How to Be Fit and Like It.[20] With data from some cases and committee judgment

[20, 21] See reference in bibliography at the end of this chapter.

on what was feasible, suggestions for norms were made to be considered useful on a national scale.

It was with this background and these needs that the teachers set about solving this problem. By the summer of 1943 when the United States Office of Education published their second manual, one on physical fitness for college students,[22] they had reached the first stage in more satisfactory testing. That stage was the analysis of physical fitness into its component parts and selection of tests for these specific qualities. This was followed almost immediately by the proceedings of the physical fitness workshop which was sponsored by the National Association of Physical Education for College Women.[15]

The approach in the two cases was similar. It seems to be the only logical one. Intelligence tests have been broken into various parts, e.g., number, verbal and deductive factors. This is exactly the same procedure. The two sources mentioned above, the college manual and the workshop proceedings, list the following elements in common: Strength, endurance, agility or body control, flexibility, and posture. The latter bulletin includes also kinesthesia, power, and speed.

The Women's Army Corps conditioning program[24] follows a similar analysis. Considerable emphasis is put on the development of strength of different muscle groups with a suggested test for each. The other items are posture, agility, balance, coordination, speed, relaxation and specific skills.

The programs of fitness testing developed in the schools show certain characteristics and trends. Much more work has been done on tests for the college level than for the high school level. Fitness tests have been used more at the college level than at the high school level. There probably are several reasons for this. It is partially a reflection of the fact that more suggestions have been made for the older group. The apparent immediacy for the college student of military or civilian work demanding a high level of fitness has resulted in greater emphasis on development of fitness in the college programs.

One of the outstanding points to be noted in studying testing programs is the great variation of test items. Some schools give a very limited battery, others a very extensive battery; but brief or

extensive there is very little exact duplication from school to school. The similarity lies in testing for specific phases, usually of "motor fitness." The batteries are as varied as the schools. For example, the Research Committees of the Central and Midwest Districts of the National Association of Physical Education for College Women conducted surveys in 1944 to determine the fitness tests in use in the various schools of those two districts. The results showed approximately fifty different tests in use for the evaluation of fitness. There seemed to be more similarity in tests used by various schools within each district than between the two districts.

An example of a testing battery adjusted for the high school girl is a study made in Illinois.[16] This study sets up minimum standards and achievement scales for twelve different items. The National Section on Women's Athletics of the American Association for Health, Physical Education and Recreation also suggested standards of performance on a number of tests.[18]

The college men's plans for testing fitness show almost as much variation as those for women. There is perhaps a little more influence from testing programs in the armed services and for that reason more agreement on the basic items. An example of the men's battery in wide use is that developed in Indiana.[1] In general, the men have attempted by means of tests to evaluate the results of the program more than has been done for the women. Examples of this work appear in the bibliography.[10, 25]

SELECTION OF FITNESS TESTS

When tests are selected for specific qualities, then a few or many may be chosen according to the kind or level of fitness for which one is striving, the time available for testing, and the merits of the various tests. The first point must be decided in each particular case when one knows what duties the students are or will be assuming. The other two points will be discussed.

Economy of time in test administration has already been discussed with respect to general programs. The possibilities for economizing in this type of testing are probably greater than in many others. The first advantage is that many tests can be given in mass. Mass tests are usually limited only by the number who can

be put into the testing area. Partners are instructed on the method of scoring and the test then requires just long enough to give it to two persons. Tests are usually limited because they require equipment or special areas for their performance. For example, arm strength may be measured by pull-ups or chinning in various forms, by push-ups in several forms, or by push-pull tests with the dynamometer. All forms of hanging and pull-ups require equipment, and that equipment is of such a nature that it is usually very limited. The dynamometers are expensive and often unobtainable. That leaves the push-up test which can be given at different levels of difficulty without equipment.

By avoiding use of two or more tests which measure the same thing, time can be saved. This can be determined by a subjective analysis of what the tests involve, or preferably by statistical analysis if such is available. For example, Wilson's[26] study on strength tests for college women indicated certain relationships in terms of correlation coefficients.

TABLE IX

INTER-CORRELATION OF STRENGTH TESTS IN THE WILSON STUDY *

Related too highly to give both tests								*Not highly related*				
	1	2	3	4	5	6	7		1	2	3	4
1. Bent arm hang								1. Bent arm hang				
2. Push-up from knees	.43							2. Weight holding	.34			
3. Push-up on bench	.54	.76						3. Push-pull	.25	.21		
4. Pull-up	.56	.68	.63					4. Vertical pull	.23	.23		
5. Pull-up (knees bent)	.54	.48	.53	.58				5. Rope climb		.12	.27	.26
6. Rope climb (arms only)	.52	.61	.62	.62	.46			6. Push-up (bench)		.37	.10	.10
7. Push-pull								7. Pull-up		.27	.27	.35
8. Vertical pull							.61	8. Pull-up (knees bent)			.30	

* Adapted from Table I, based on 52 subjects.

The Wilson study started with tests which were believed to measure similar capacity. Statistical evidence bears out similarity in some, differences in others. Mohr's[14] study started with tests which were believed to be different as judged subjectively. The statistical evidence corroborates that selection.

TABLE X

INTER-CORRELATION OF TESTS IN THE MOHR STUDY *

	1	2	3	4	5
1. Sit-ups					
2. Chair stepping	.101				
3. Push-up	.494	.043			
4. Bouncing	.261	.218	.329		
5. Pull	.229	.101	.042	.207	
6. Obstacle race	.411	.077	.349	.341	.262

* From Table I, based on 140 cases.

The scoring method may affect the time required. Most of these tests allow only one or two trials so there is no problem of many repetitions. However, a test may be set up with a specified time limit and scored on the number of repetitions performed in that interval, or it may be scored as the time required to perform a specified number of movements. The uniform time is better as it makes it possible to administer to the whole group with a single timer and stop watch, to know the exact time necessary, and to avoid having the faster ones wait while the slow ones perform. If the duration of the test is comparatively short in both cases, the capacity measured is essentially the same, that for maintaining a maximum rate for a reasonable length of time.

Tests should grow directly out of regular instruction. In that case a minimum of specific instruction and special practice will be necessary when testing is conducted. Then with the careful organization of partners or assistants, and a scoring and recording system, little time need be spent in testing. Building of fitness requires time; most of us have too little time to accomplish what we wish. Let us not waste *unnecessary* amounts on testing.

Additional time will be saved if the tests are self-administering or can be given to each other. This means that the scoring must be simplified and the performance so obvious that no errors in scoring can be made after brief instructions. When each person must be judged by the teacher, or equipment must be adjusted for each person by assistants, too much time is spent in waiting. Furthermore, the tests must be given to all in sequence rather than simultaneously.

Tests which require equipment of any type create a problem. Strength testing equipment is not on the market. Where spring balance scales can be obtained they may be used for substitute tests very successfully. They have been found to be about as valid and reliable as the dynamometer tests.* If equipment seems necessary, use that which is already in the gymnasium. Obstacle courses have been used quite extensively as a means of measuring endurance and general agility. Most of these require considerable space and labor to set them up. Shorter courses may be arranged in the gymnasium with equipment which is there. In any case in which the obstacle course is used, instruction should be given on each part of it and practice permitted before it is done under pressure of time. This will do much to avoid injuries. The obstacle courses in schools have frequently been built for the boys and then used by the girls. This invites injuries because the distance to be jumped or the heights to be scaled are usually too great for the majority of the girls. The boys' obstacles are seldom padded sufficiently to give adequate protection for the girls. This problem may be partially solved by providing different obstacles along the same course for the girls.

Another requirement of good tests is that there be no zero scores, or no undue massing of scores at any point along the scale. This is sometimes very difficult to avoid in this type of testing because of the great range in ability and because the standards we consider desirable are apt to be far above the ability of many of the group. One might argue that the tests should not be modified for the weaker group because it gives them a false impression of their status. However, the purpose of the whole testing procedure, motivation is defeated if some adjustment is not made. If a student and all his friends in the class get a zero score he is much more impressed by that fact than he is that a few persons made a very high score. Tests may be modified to suit this large group of less capable ones and then increased in difficulty as they improve. In other words

* Wilson, Marjorie, *ibid.*
Reliability—grips, pull, push, .94, .89, .76 respectively; spring scale: thrust, horizontal pull, vertical pull, .91, .82, .91 respectively. Validated against Rogers short strength index, push and pull, .49; vertical pull, .59
Mohr, Dorothy, *ibid.*
Reliability—vertical pull, .93

the tests should be adjusted so as to be a challenge, rather than a source of discouragement and criticism. If this rule is followed it will eliminate some forms of the push-up, of the sit-up, or climbing for most classes.

In order to obtain this progression in difficulty, and to select tests wisely on the basis of abilities involved, a careful analysis of each test must be made. For example, a sit-up is much more difficult if the legs are not held down, and much easier if the head and arms may be allowed to lead. Likewise, a running test may measure speed if set up to get maximum performance for a very short time, or it may measure endurance if planned for a longer period at maximum or submaximal speed. Another running test may measure leg strength, and in slightly different form, agility. If tests are to be used in a series this analysis of the test is essential for proper sequence of tests.

If tests are to measure ability accurately, they must result in little or no muscular soreness afterward and a limited amount of discomfort while doing them. Motivation is the controlling factor here which will encourage effort in spite of consequences, but effort can not be forced, especially against such odds. College women are much more apt to consider such discomfort unnecessary, and the person who works to that point a "sucker," than will high school girls. Hence, this point is much more important in working with college women. The student may work to the limit the first time but she remembers her previous experience and will not really try the second time. There are certain tests which students associate with immediate or later discomfort and hence they develop a distinct dislike for the test, a dislike which is apt to become apparent throughout the group. Soreness can be avoided to a considerable extent by aiming the conditioning program directly toward the tests. The WAC program has followed this principle very well in the organization of their physical training system, and in practice it has apparently succeeded in this respect. However, their tests are performed for a certain number of times, usually limited to a rather conservative number of repetitions or limited time interval. That is an additional factor conductive to better results. The physiological effects on a muscle worked to the limit can not be totally avoided even in a trained muscle. Fatigue products accumulate and

the muscle works under ever less favorable circumstances as the muscle approaches the contracture of fatigue. The protective reaction of the organism tends to make the person decrease effort when it becomes too painful; it therefore takes a *super* type of motivation to get continued effort on endurance tests after discomfort sets in. Not only is submaximal effort obtained for practically all, but they quit at very different points along the fatigue curve. This makes the measure still more unreliable.

Safety should be considered as carefully as any quality in selecting tests. As in any activity taught there may be an occasional injury, but most of these can be prevented by forethought. Certain tests such as sit-ups or squat thrusts, are conducive to strains in the abdominal muscles, or various ill effects on viscera by increased abdominal pressure. These certainly should not be engaged in by anyone subject to conditions such as hernia, appendicitis, or severe dysmennorhea. Likewise, the squat thrusts, deep knee bends, and many forms of jumping are hard on knees, particularly if the knee is already weak, or if the person is extremely overweight.

Some tests of endurance involve a run of 150 to 200 yards or more. Such runs should not be taken unless the individual has been trained for such strenuous activity. Standards of competition for women and girls have always contra-indicated such events. Tests are essentially competition, competition against standards, or the rest of the class, or one's previous record. It would seem that similar criteria concerning strenuousness should be applied to the so-called competitive events and to test events. The minimum requirement certainly should be adequate training.

Adequate training to insure safety in the test suggests another practice which should be avoided. Tests are frequently given very early in a season or semester. The best means of showing improvement is to test at the beginning and again at the end of the learning interval. Care should be taken, particularly on endurance events, that they are not given without reasonable pre-conditioning.

If obstacle courses are to be used for junior and senior high school ages, provisions should be made for adjustable or alternate obstacles for the very small child. Obstacles which are to be climbed or which the person must drag himself over are not very desirable

for girls from the safety standpoint, and if used should be well padded.

Before a teacher starts a program of testing for fitness it seems important that certain questions be considered. The program is largely determined by the answers.

(1) What aspects of fitness am I teaching?

(2) Is that kind of fitness measurable, and if so in what units?

(3) Shall I attempt to measure comprehensively or just one or two aspects?

The answers to the above questions would doubtless point toward the use of comparatively few tests of the qualities of total fitness which are best achieved through physical education experiences. The questioning then continues.

(4) Shall I rely only upon highly refined measures or something which in spite of technical shortcomings will motivate interest and effort?

(5) What measures are available which will have the least technical shortcomings?

The decisions on these two questions may leave one only a very limited choice or may give one opportunity for experimentation and ingenuity. Both courses have advantages. The final questions may be of this type.

(6) Is time of training sufficient to produce results which are measurable?

(7) Is time which is available sufficient to justify more than very brief testing?

There is similarity between the fitness tests in use for the two sexes. However, some variations are essential. The tests suggested in this chapter are primarily for the girls and women. A few of the tests may be adapted for boys, or other tests may be selected in keeping with the considerations discussed above.

In most of the classifications below, two or three tests are suggested. Seldom, if ever, would that many tests for a given purpose be used for a single class. Careful selection should be made from these suggestions, or from other tests, according to the ability of the class, the equipment available, and the requirements for good testing suggested above. These tests have been found valuable but others may be better adapted to certain situations.

In choosing a battery of tests some consideration should be given to their cumulative effect on the students, especially if they must be given in a single class period. When a series of extremely strenuous tests is given with the tests following in rapid succession, there may be fatigue to the point where the scores are reduced, and the students are excessively fatigued. Alternation of types of tests during administration will be helpful, but if a very extensive battery is to be given it is usually preferable to give by halves on successive days.

TESTS OF STRENGTH

Tests for Arm and Shoulder Girdle

1. PUSH-UP (ON KNEES)

Description: Lie in the prone position with hands under the point of the shoulders, elbows spread a little (*a* in Figure 25). Feet may be raised from the floor or not as preferred. Keep body straight and extend arms fully; weight will be resting on hands and knees (see *b*). Bend arms so that chest again touches the floor. Repeat

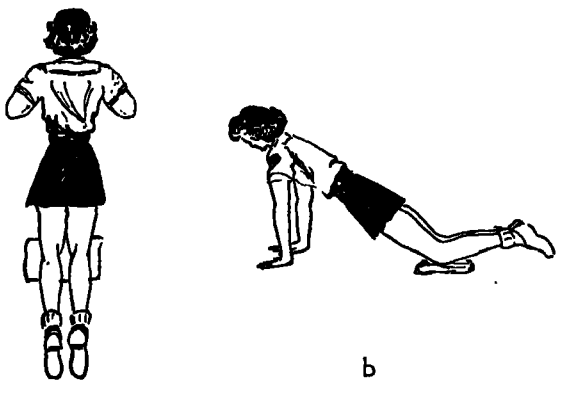


Figure 25. Push-up from Knees

a — top view b — side view

promptly and continue as long as possible or for a stated number of times. Score is the number which can be done before stopping or before position is changed.

Suggestions: If the hips and lower back are allowed to sag at all

the weight is transferred progressively up the thigh, hips, and trunk during body lowering, and in reverse order in the lift. It is necessary to keep the back straight and to hold the hips in very slight flexion (not exceeding 5°) in order to keep the weight on the arms.

If the performer keeps up regular breathing it will be more comfortable and tend to give greater endurance.

The knees should be protected by placing performers on mats, or if the floor is not too slippery by folding a sweat shirt or towel under the knees.

Train students in proper form and in judging form before giving the test. Then give as a mass test to half the class with partners scoring.

2. PULL-UP (ON HORIZONTAL BAR)

Adjust the bar to the level of the xiphoid (angle between the ribs at the base of the sternum) when the subject stands erect.

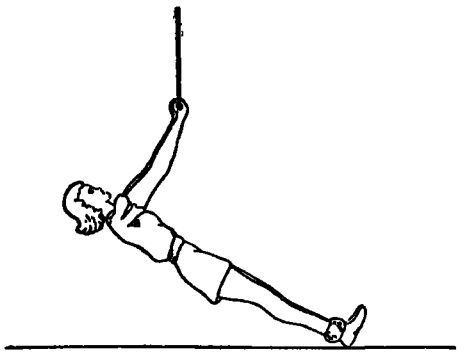


Figure 26. Pull-up on Horizontal Bar

Description: Grasp the bar with the hands about shoulder width apart and palms toward the face. Move the feet far enough beyond the bar that when the weight rests on the heels, with knees, hips and back straight, the line of the body forms a right angle with the line of the straight arms. (See Figure 26.) Keep the body straight and bend the arms until the neck or upper chest touches the bar. Extend arms again. Repeat again without pausing and continue as long as possible or for a stated number of times. Score is the number which can be done before stopping or before the body begins to sag or sway.

Suggestions: The sides of parallel bars may be used if a horizontal bar is not available. This does not permit quite as exact adjustment for height. Flying rings may also be used but give less uniform results because they swing.

Usually equipment is very limited so only a few need be trained to administer the test.

A similar form may be done with a partner who holds a wand and stands astride the performer who starts lying on her back. The performer grasps the wand and pulls up until the chest touches the bar, lowers to full arm length, and repeats. The top person must stand straight and firm, and with feet far enough apart to permit the performer to reach the wand. Sometimes the top person tires before the one who is taking the test. For that reason it seems better to have two persons hold the wand, one at each end. One of the advantages of these forms is that there is no need for adjustment of the bar. If no special equipment is available, small size softball bats may be used.

Another form of this test puts the feet flat on the floor, knees are bent at right angles, trunk and thighs are horizontal, and arms are vertical. When the chinning is done the movement is at the knees, not at the ankles. A little practice is necessary to standardize form.

3. PUSH AND PULL

Use the hand dynamometer with push-pull attachment.

Description: Hold the apparatus in front of the chest, one hand on each handle; elbows are bent and arms in a horizontal plane. Pull as hard as possible. This pull is similar to that exercise in which one tries to touch elbows behind the shoulders. (Assistant records score and resets dynamometer.) In same position push in on apparatus as hard as possible; the heel of the hand may be used. Do not brace the apparatus against the chest in either trial.

Suggestions: If the handles have sharp edges they should be covered or padded with adhesive tape or something that will not slip.

4. VERTICAL PULL (WITH SPRING SCALE)[26]

Use a good grade spring scale. The lower end of the scale is fastened to the floor. A rope, securely fastened to the upper end

of the scale, is run through an overhead pulley so that the handle end can be reached from a standing position, with the arm slightly flexed. The length of the rope should be adjusted to the height of the shortest girl. The rope may be shortened for taller girls by slipping a wooden peg through one or more loops of the rope.

Description: Stand erect, in a comfortable stance and with shoulders fixed, pull down as hard as possible without bending the knees or hips, or twisting the body. Score is the number of pounds registered on the scale.

Suggestions: The examiner must squat to the level of the scale in order to read it accurately.

Watch to see that the subject does not bend knees or trunk, then quickly read the scale while she continues pulling.

Take the girls according to height in order to save time in adjusting the rope. Give instructions to all rather than individually as each starts the test.

TESTS OF ABDOMINAL STRENGTH

1. SIT-UP

Description: Lie on the back with knees bent, partner holding the feet down firmly. Fold the arms and hold against the chest. Rise to an erect sitting position, return to back lying. Repeat as many times as possible or for a stated number of times. Score is the number of times the erect position is assumed.

Suggestions: It is permissible to lead a little with the head, and this makes it unnecessary to judge for fouls. The elbows are kept down to prevent the momentum of an arm swing.

Variations may be used; advantages and disadvantages follow:

(1) hands behind neck, elbows kept back—encourages forward head and permits excessive arm pull

(2) hands on the shoulders, elbows against the ribs—reduces arm momentum, but puts the weight of the arms forward and therefore makes the test easier

(3) arms extended along sides and hands kept on the floor—avoids a high arm lead, but may permit a push-off with the hands or elbows from the floor

(4) arms extended, hands sliding along on top of thighs—difficult

to prevent the subject from pulling up by grasping the thighs or clothing

(5) hands on shoulders, legs straight, bend forward far enough to touch the elbow to the opposite knee, alternating right and left —uses oblique abdominal muscles more in latter stage of the sit-up, encourages pronounced spinal flexion

(6) head, neck and back held straight throughout—good as an advanced and difficult form, difficult to judge and score

(7) no support on feet—makes the test very difficult, may result in many zero or low scores

(8) legs straight, heels on floor—there is a tendency to hollow the lower spine excessively, the hip flexors do more of the work than when the knees are bent

2. ROCKER

Description: Start in back lying position, hands on the shoulders with elbows extended straight sideward. Raise the feet and trunk off the floor a few inches so as to be supported only on the hips. Rock gently from side to side while holding this position. Sideward movement is checked by contact of the elbow on the floor, a push-off with the elbow is permitted. Score is the number of seconds the position and action are maintained.

Suggestions: Continue regular breathing for greater comfort.

Variations are sometimes used:

(1) V position held with no rocking movement, arms extended just off the floor. Rocking gives the advantage of momentarily, moderately relaxing the muscles on alternate sides of the abdominal wall; therefore, it slightly prolongs the time and gives better distribution of scores. In the V position, the extended arms are hard to keep off the floor or thighs.

(2) V position, held with hands holding the thighs. The position is fixed and the abdominal muscles are almost entirely relieved of action. It becomes primarily a balance test.

Test of Foot Strength

1. BOUNCING

Description: Take a full squat position with the knees fully flexed and clasp the arms around the legs in such a way that the

knees can not extend. Bounce continuously in place by extending the ankles, using enough force to come just off the floor. Continue as long as possible. Score is the number of bounces which can be made at the rate of 120 per minute. If balance is lost or arms unclasped, resume position and continue, providing not more than 3 counts are lost. Deduct the number lost from the total count.

Suggestions: Cross the arms and grasp near the elbows, otherwise the knees and hips extend. All force should come from the ankles and feet. If bouncing is low there is less tendency to try to use the knees.

Counting should be done centrally. Use a metronome or drum to keep the rhythm. Count so all may hear and the student notes the count on which she quits, subtracting any lost counts before recording her score. Partners may do the scoring.

TESTS OF ENDURANCE

Endurance is the most difficult aspect of fitness to measure. Endurance is primarily the result of a physiological capacity of the organism to continue functioning satisfactorily. Endurance may be either the ability to maintain action at maximum speed for a short period of time, or the ability to maintain action at a slower rate for an indefinite period of time. The former type of endurance is the easiest on which to set up tests, but it does assume that the subject will put forth the effort to work at a maximum rate.

The tests listed here are considered the best of those now in use, when the criteria presented in the first of the chapter are applied. Since cardio-respiratory function is indicative of general organic endurance, such tests are frequently used. Only one such test is listed here since they are not practical for most physical education teachers. The other tests aim at measuring endurance to perform at a high rate for a comparatively short period.

The first test may be given to one-third the class at a time. The latter two are limited, usually to one or two students at a time, because of equipment required, or the nature of the test. Choose *one*.

1. CHAIR CLIMBING

Use ordinary chairs; arm chairs from a class room are suitable. The only requirement is that all are of the same height for any groups to be compared. The number required is one-third that of the students to be tested.

Work in groups of threes. *One* performs, *two* holds the chair and counts aloud the number of times *one* mounts the chair. *Three* holds the chair and writes the scores on an individual score card. *One* stands in a position of readiness beside the chair with one foot on the chair, and holding *two's* left hand with her own right hand. On the signal, Ready, Go! she rises to an erect standing position with both feet on the chair. Immediately she steps down to the floor with the same foot which she started on the floor. She continues as rapidly as possible until the final whistle. She may change feet occasionally by making the shift while both feet are on the chair. She may hold on to *two's* hand throughout the test.

The timing is done centrally by one timekeeper, who has a stop watch and two whistles with distinctly different sounds. (A timer's horn works very well for one.) The timer gives the starting signal, at the end of 5 seconds she blows a whistle, again at the end of 20 seconds she blows the same whistle, then at the end of 50 seconds she blows the second whistle which is the final signal. *Two* counts continuously from "Go" to the final whistle. *Three* listens for all three whistles and immediately after each one writes the count she heard simultaneously with or just preceding the whistle. The three then change places and the test is repeated twice so each one performs.

The endurance score is the ratio between the score for the 15 second interval and the score for the 45 second interval. These are obtained by subtracting the count for the first 5 seconds from that at 20 seconds and at 50 seconds. The 45 second score is then divided by that for the 15 seconds. If the 15 second score is less than 16, then 16 is used as the divisor.

Suggestions: The assistants should hold the chair firmly to prevent its slipping. *One* grasps *two's* hand as a safety measure to prevent loss of balance or unsteadiness when she becomes tired.

The performer can postpone the fatigue of the leg muscles if she alternates frequently the foot left on the chair. She will save time

the results. However, the heart rate is so variable and dependent upon so many factors that the value of this test would seem to be reduced as much or more by occasional lack of effort.

TESTS OF AGILITY

1. OBSTACLE RACE

The race which is described in Chapter 6, p. 136, may be used as an agility test, particularly after practice has been allowed on the race as a whole, or on its parts. If given at the beginning of the year, a single administration may serve the purpose in both batteries.

2. SHUTTLE RACE

This race is described on p. 149, and may also be used for this purpose. It will be a little more economical of time than the obstacle race since it can be given to several at the same time.

TESTS OF FLEXIBILITY

Flexibility must be considered in terms of a given joint or of adjacent joints, just as strength is considered with reference to specific muscle groups. Flexibility is desirable only as it contributes toward some other ability or to freer movement. Determine the type or types of flexibility desired and choose one test for each type selected.

TESTS OF HIP AND BACK FLEXION

1. STANDING, BOBBING

Arrange a scale marked in half inch intervals on the front of a chair, stool, or platform with the inches marked above and below the level of the chair surface. The scale should be not more than 3 to 4 inches wide, and the chair must be stable.

Description: Stand with toes even with the front edge of the chair and against the sides of the scale. Let the arms and trunk relax and hang forward, fingers in front of the scale. Then bob

downward forcefully three or four times reaching equally with the fingers of both hands. The knees must be kept straight. Score is the lowest point reached in the series of bobbings.

Suggestions: A 20-inch scale may be marked from top downward attached so that 10 represents the level of the top of the chair. This usually provides measurement for the range of flexibility found in a class. If the suggested standard of performance is for some point above or below the chair level this is the best arrangement.

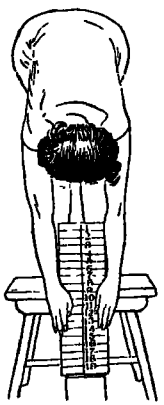


Figure 27. Standing Bobbing Test

The other alternative is a similar length scale, but the chair level is zero with deviations progressing upward or downward from that point. The score is then minus if the reach is short of the stool level, or plus if it is beyond that level. If the accepted standard is the ability to reach the surface on which one stands, this is a very descriptive form of scoring and is preferable.

The test may be scored as the reach which the performer can attain and hold for two or three seconds. This will slightly reduce the scores, and will make it easier to score.

In order to score accurately, the person giving the test must have the eyes down on a level with the reach. A student assistant is helpful. She can watch for bent knees and also be ready to give support if the performer loses her balance.

2. SITTING, BOBBING

Description: Sit on the floor with the legs straight and extended at right angles to the line of the boards in the floor. Place the heels on a crack between the boards and with feet about five or six inches apart (just wide enough to get the hands between the heels). A partner stands with her feet against those of the performer to prevent the latter from slipping. The performer bends forward reaching as far forward as possible on the floor. The knees must be kept straight. Score is the farthest point reached by the finger tips.

Suggestions: This may be given as a mass test and since its main advantage over standing bobbing is for quickness of administration it would seem advisable not to use special floor markings. If sufficient rulers are available they may be held in place by one heel of the standing partner. The zero point is then the heel line and the score is plus if the reach is beyond the heels or minus if short of them.

The most satisfactory means is by using the boards of the floor as units of measurement. Most gymnasium floors are laid with 2¼ inch boards. The score is measured to the nearest half of the board (in other words to the nearest inch). For example, if the reach is almost to the middle of the second board beyond the heel, the score is + 1½. If the reach is almost to the second crack beyond the heel the score is + 2. If narrower boards are used, count to the nearest crack.

A variation of these which is sometimes used is done in a sitting position with the legs straight and together. Bend forward to touch the forehead to the knees. If the head touches, spread the legs just enough to get the head between the legs. Bend forward to touch the forehead to the floor. This method has the advantage of eliminating all direct measurements and of reducing the influence of arm length. It also includes neck flexion which is omitted in the other forms. However, it leaves most of the class in one of two groups, those touching the knees, and those not touching. There will be a third, comparatively small group who reach the floor. This may be adequate differentiation and is faster, each one scoring her own performance.

These tests all measure almost identical forms of flexibility. How-

ever, the standing scores will always run a little higher because of the more effective use of gravity and because the hips are shifted back of the heels when standing, thus shortening the distance to the feet. They are measures not only of back and hip joint flexibility but also of elasticity and relaxation of the hamstring (posterior thigh) muscles. This is doubtless of more importance as a basis for success in certain activities such as dance or tumbling rather than in every day activities. It also seems to be related to ease and economy of muscular effort in many movements.

Students who take flexibility tests sometimes comment on the disadvantage which they have because of the shortness of legs or arms. Teachers who are considering the use of the tests are also prone to think of body build as a determining factor in the scores obtained. With this in mind anthropometric measures were taken on a hundred college women acting as subjects in a flexibility study.* There is no evidence from these data to indicate that variations of body build as found in an average group would affect flexibility scores unduly. The correlations with the standing bobbing test (p. 128) and the measures which might appear to affect bobbing ability follow:

Height	+ .157
Trunk length	+ .148
Arm length	+ .294
Trunk and arm length	+ .297
Ankle flexion	+ .178
Spinal extension	+ .262

The correlations between spinal extension (described in the following test) and various measures are equally low.

Shoulder flexibility	+ .262
Trunk length	+ .180
Pull strength	+ .138

Test of Extension in Upper Back

1. SPINAL EXTENSION

Description: Lie in a straight prone position on the floor or a table. Hands are clasped together above the hips. Raise the head and shoulders from the floor by arching the upper back; pull with

* Wilson, Marjorie, and Scott, M. Gladys: A Study of Flexibility in Relation to Physical Education Activities, unpublished study.

the arms keeping the lower corner of the ribs on the floor. Score is the vertical distance from the suprasternal notch (top of the sternum) to the floor.

Suggestions: Fixation of the ribs on the floor is best assured by an assistant who places a finger on the lower points of the rib cage where contact is to be maintained. The assistant can ask for adjustment of position if she arches so much in the lower back that the ribs leave the floor.

Measurement can be made most easily by placing one end of a string on the suprasternal notch when she starts to lift. The string is pulled taut and straight to the floor (vertically) while at the top of her extension. Measurement of the string from finger tip to finger tip is then made on a ruler. The score is read to a scorer.

Measurement is most rapid if students lie down side by side, always with one or two ready ahead of the one being measured. When each girl is measured she may go on to the next test or to practice.

This type of flexibility facilitates many movements. A reasonable degree is also important in maintenance of good posture. To a slight extent it also measures strength of spinal extensors since the body must be lifted and held there momentarily for measurement.

Range in trunk twisting is also desirable in many activities but measurement techniques are highly unreliable. Measurement should be in a sitting position if movement is to be limited to the spine, standing position if hip rotation is to be included. Variations in shoulder girdle action also add to the unreliability, and variability from one individual to another.

BIBLIOGRAPHY

1. Bookwalter, Karl W.: "Test Manual for Indiana University Motor Fitness Indices for High School and College Men," *Research Quarterly,* 14, December 1943, p. 356
2. Brouha, Lucien: "The Step Test: A Simple Method of Measuring Physical Fitness for Muscular Work in Young Men," *Research Quarterly,* 14, March 1943, p. 31
3. Brouha, Lucien, and Gallagher, J. Roswell: "A Functional Fitness Test for High School Girls," *Journal of Health and Physical Education,* 14, December 1943, p. 517
4. Cassidy, Rosalind, and Kozman, Hilda Clute: *Fitness First, A Physical Fitness Workbook for High School Girls.* A. S. Barnes & Company, New York, 1943

5. Clarke, Harriet: "A Functional Physical Fitness Test for College Women," *Journal of Health and Physical Education*, 14, September 1943, p. 358
6. Cureton, T. K.: *Physical Fitness Workbook*. Stipes Publishing Company, Champaign, Illinois, 1944
7. Espenschade, Anna: "Report of the Test Committee of the Western Society of Departments of Physical Education for Women in Colleges and Universities," *Research Quarterly*, 14, December 1943, p. 397
8. Hall, D. M., and Wittenborn, J. R.: "Motor Fitness Tests for Farm Boys," *Research Quarterly*, 13, December 1942, p. 432
9. Havlicek, Frank J.: "Speed Sit-ups," *Research Quarterly*, 15, March 1944, p. 75
10. Kistler, Joy: "A Study of the Results of Eight Weeks of Participation in a University Physical Fitness Program for Men," *Research Quarterly*, 15, March 1944, p. 23
11. Leighton, Jack R.: "A Simple Objective and Reliable Measure of Flexibility," *Research Quarterly*, 13, May 1942, p. 205
12. Miller, Ben W., Bookwalter, Karl W., Schlafer, George E.: *Physical Fitness for Boys*. A. S. Barnes & Company, 1943, Chapter X
13. Minnesota State Defense Council: "Physical Fitness and Recreation Program," *Bulletin No. 2*, Series B, St. Paul, Minnesota
14. Mohr, Dorothy: *Measured Effects of Physical Education Activities on Certain Aspects of the Physical Fitness of College Women, Research Quarterly*, 15, December 1944, p. 340
15. National Association of Physical Education for College Women: *Proceedings of Victory through Fitness Workshop*, 1943, pp. 43-48
16. O'Connor, Mary Evangeline: *Motor Fitness Standards for High School Girls*. Unpublished M.A. thesis, University of Illinois, Urbana, 1944
17. Phillips, Marjorie, Ridder, Eloise, and Yeakel, Helen: "Further Data on the Pulse Ratio Test," *Research Quarterly*, 14, December 1943, p. 425
18. Research Committee, National Section on Women's Athletics: "Physical Performance Levels for High School Girls," *Journal of Health and Physical Education*, 14, October 1943, p. 424
19. Springfield College Studies: "Physical Fitness," *Research Quarterly*, 12, Supplement May 1941
20. Steinhaus, Arthur H, Hawkins, Alma M., Giaque, Charles D., Thomas, Edward C.: *How to Keep Fit and Like It*. George Williams College, Chicago, 1943
21. United States Office of Education: "Physical Fitness through Physical Education, Victory Corps Series," *Pamphlet No. 2*. Superintendent of Documents, U. S. Government Printing Office, Washington, D. C., 1943, p. 81
22. United States Office of Education: "Physical Fitness for Students in Colleges and Universities." Superintendent of Documents, U. S. Government Printing Office, Washington, D. C., 1943
23. United States Office of Education: "Scales for Tests for High School Boys of Strength of the Abdomen and Back," *Education for Victory*, 3, August 21, 1944, p. 4
24. United States War Department: *WAC Field Manual*, 35-20; *Physical Training*. Superintendent of Documents, U. S. Government Printing Office, Washington, D. C., 1943
25. Wilbur, Ernest A.: "A Comparative Study of Physical Fitness Indices as Measured by Two Programs of Physical Education: The Sports Method and the Apparatus Method," *Research Quarterly*, 14, October 1943, p. 326
26. Wilson, Marjorie: "A Study of Arm and Shoulder Girdle Strength of College Women in Selected Tests," *Research Quarterly*, 15, October 1944, p. 258.

6.

Measurement of General Motor Ability

MOTOR ABILITY DEFINED

Every teacher knows that some students learn very much more rapidly and with less apparent effort than other students. It is also well known that after an interval of instruction and practice some students have a greater variety of skills and greater proficiency in them. This is frequently explained rather vaguely as a difference in motor ability. This is actually true, and since it affects the teaching-learning situation so deeply it seems worthy of careful consideration.

Writings by various authors have led to the use of several similar terms with different connotations. This has sometimes been confusing; therefore, discussion of these terms follows.

Motor educability is the inherent aptitude (motor and mental) for learning new skills quickly and effectively. In carrying out this concept it is understandable that tests of this characteristic should involve motor problems new to the subject; that they be presented through the usual media of instruction, verbal description and demonstration; that they prohibit preliminary practice, and allow very few trials; that they be of a success or failure type. Therefore, most of the tests which are proposed for measuring educability are of a stunt type, and usually include several stunts in order to secure satisfactory reliability. The greatest difficulty is in devising motor problems which are new to the performer.

Motor capacity is very similar to educability but is really a little broader. Such batteries may contain an educability test plus general agility items such as obstacle, or dodge runs, or Burpee tests.

Physical capacity implies a fitness or capability for performing motor activity. Since that capability is dependent upon several

things the tests vary from physiological functioning of circulatory and respiratory systems, to strength, or reaction time. Such tests will not be considered here as they are partially taken care of by proper medical examination, and partially in fitness testing.

The term *motor ability* is sometimes used to mean achievement in basic motor skills; or it may be interpreted as a more general term combining the concepts of motor educability and achievement. How successfully achievement and educability can be separated is still an unsolved question. Motor ability measurement is usually concerned with some form of running, throwing, and jumping; tests are repeated from time to time, and practice on them is permitted. The level of ability recorded may be due to capacity for neuro-muscular coordination, to practice, to strength, or other less evident factors.

It seems that the information needed by the teacher concerning the student, at least from the junior high school age up, is aptitude for learning, ability in the fundamental skills, and ability in the various sports or activities. The first two points are of a general nature and can be interpreted in relation to any activity. Further ability is specific for each activity. In every case it is necessary to abbreviate testing as much as is possible and to relate them whenever possible.

REQUIREMENTS FOR A MOTOR ABILITY TEST

Since it is very difficult to separate the measurement of aptitude and of achievement, and for the sake of abbreviation of the testing procedures, it would seem advisable to follow the second interpretation of motor ability stated above and consider them as dual and interdependent aspects of *general motor ability*. Let us consider the requirements for such a set of measures.

(1) First of all it would be necessary to have unusual situations, or motor acts relatively new to the subjects.

(2) Students should not practice on the test as such.

(3) It is essential that students be given a clear idea of the problem presented by the test but that should not include specific coaching or instruction on techniques to be used.

(4) Principal activities in the physical education program should be analyzed for the skills that they have in common, for example, balance and weight control, eye-hand coordination, strength, agility, speed, etc. The tests should be set up to include as many of these as possible.

(5) Tests combining more than one element in a significant way should be used when possible.

(6) Part of the test should give opportunity to demonstrate skill developed by those who have worked hard previously.

(7) The tests should not put undue emphasis on endurance, strength or any other one factor.

A MOTOR ABILITY BATTERY

The motor ability test battery presented here has been successfully used both with college women and high school girls.[12] Let us consider this battery in light of the above criteria which are peculiar to this type of test and supplement those criteria discussed in Chapter Two.

The minimum battery recommended is the obstacle race, basketball throw and standing broad jump. The 4-second dash and wall pass may be added or substituted for the obstacle race.

1. Obstacle Race[13]

The space needed is 55 feet by 12 feet; equipment needed, three jump standards and a cross bar at least 6 feet long; lines on the floor. (See Figure 28.)

Description: Subject starts in a back lying position on the floor with the heels at line *a*. On the signal, Ready, Go! get up and start running toward J, as you come to each square on the floor step on it with both feet. Run twice around J, turn back to *d*, go under the cross bar, get up on the other side, run to line *c* and continue running between lines *b* and *c* until you come to *c* for the third time. Score is the number of seconds (to the nearest .1 second) that is required to run the course.

[12] See references in bibliography at end of this chapter for report on college use. For high school use: Smalley, Jeannette, and Scott, M. Gladys: Motor Ability Tests in Junior and Senior High School, unpublished study.

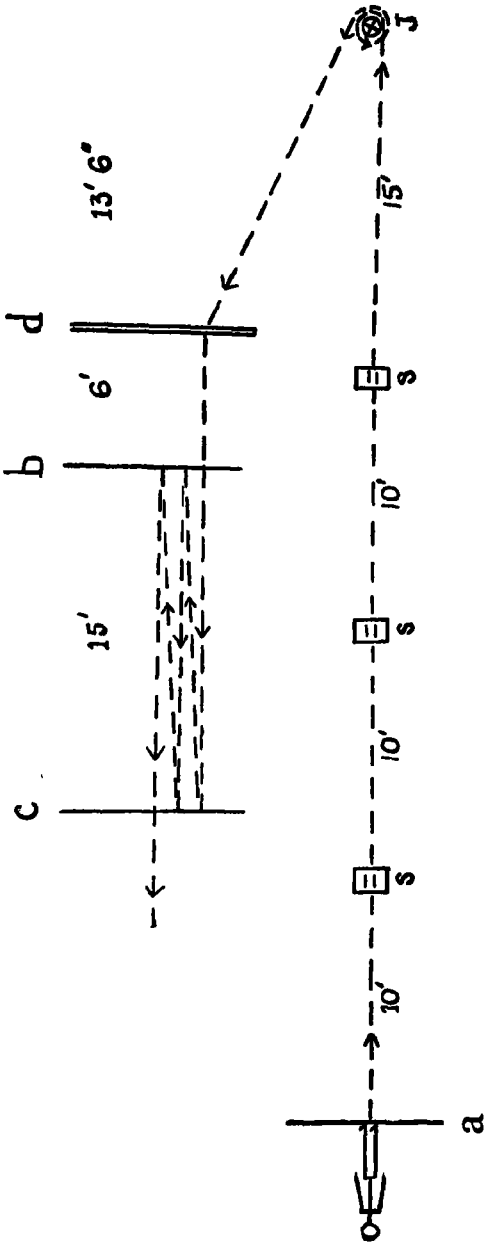


Figure 28. Floor Markings and Pathway for Obstacle Race

a — starting line
b — line for the shuttle
c — finish line
d — cross-bar (18" high)
J — jump standard
S — spot on floor (12" x 18")
– – – path of runner
distance from end of cross-bar to line of inner sides of spots, 4' 4"

Suggestions: Give instructions to all so they need not be repeated when individuals are ready to run. Demonstrate what is meant by stepping with both feet on each square.

Each successive runner should lie down as soon as the girl ahead is up. This avoids delay in starting new runners.

If two timers and watches are available, the next girl starts as soon as the one ahead finishes circling the standard. Approximately twice the number can be scored on the same course with this arrangement.

Do not call the runner back if the toe or heel extends outside of the square. Some feet are too large to fit inside the square if the heel is lowered. Judge on whether the stride is adjusted to contact the square and whether there is a transfer of weight in the square.

2. Basketball Throw for Distance

Space needed is about 80 feet long and 20 feet wide, a throwing line marked about 8 feet from one end of the course and parallel lines every 5 feet beginning 15 feet in front of the throwing line.

Description: Start anywhere you wish behind the throwing line, but do not step on or across the line when throwing. Throw in any way you wish, three consecutive times. The score is the distance from the throwing line which the ball travels before touching the floor. Only the longest throw counts.

Suggestions: Explain carefully but do not demonstrate. Answer questions about the test except those on throwing technique. If asked whether the throw should be overhand or underhand, whether from a stationary position or with a step or run, simply reply that the throw may be of any type, providing the feet are kept behind the line; the purpose is to throw the ball as far as possible. This may not be good teaching procedure but it is essential for this form of testing if you wish to know how the player is apt to meet similar problems of throwing in a game.

It is true that some will profit more than others from seeing other students perform, but they are also the ones who learn quickly from class instruction. The ones who do not profit from errors and success of classmates, doubtless will be slow to profit from class instruction.

If the gymnasium is too short and the test can not be given outside, a diagonal course across the gymnasium may be used. This insures sufficient distance in practically any gymnasium but leaves little space in which to carry on other class activities during the test.

3. STANDING BROAD JUMP

If given outside it is necessary to have a jumping pit with sunken take-off board within 30 inches of the edge of the pit. If given indoors it requires mats at least 7½ feet long and a solid board at least 2 feet long (beat boards used with apparatus are excellent) placed against the wall to prevent slipping. If the mat is marked in 2-inch intervals it eliminates the need to measure each jump with a tape.

Description: Performer stands on the take-off board, toes may be curled over the edge of the board. The take-off is from both feet simultaneously, the jump is as far forward on the mat as possible. The score is the distance from the edge of the take-off board to the nearest heel (or to the nearest part of the body if the balance is lost). The best of three trials will be counted.

Suggestions: Preliminary swinging of arms and flexing of knees are permissible providing the feet are kept in place on the board until the actual take-off.

Be sure performer understands what is to be done.

When the use of a take-off board is not feasible, jumping may be done from the mat if the mat is heavy enough that it will not slip.

The following two tests may be added if time permits the administration of the additional items. The obstacle race need not be given if these are added.

4. WALL PASS

A flat wall space is necessary at least 8 feet square. A line is drawn on the floor parallel to the wall and 9 feet from it. Several such spaces are desirable in order to test several persons at one time. Timing for all areas may be done by a single timer.

Description: Stand facing the wall, behind the line. Throw the ball against the wall, catch it when it comes back and repeat again as quickly as possible. Stay behind the line all the time. The throw may be of any type and the score is the number of hits on the wall in the time allowed (15 seconds).

Suggestions: The test may be administered very quickly if several testing areas are available and about four players start at each area. The first one is tested while the second counts the hits and watches the foul line. Then the first reports the score while the second is tested and the third counts. This is repeated until all are finished. A player who steps across the line slightly should be called back by the player scoring the trials. If the feet are in the proper position for the next throw, the error is not considered. If fouls are continuous the entire trial must be repeated.

Allow each person time for three or four practice throws before taking the test.

If a ball drops between the wall and the line it may be necessary for the player to cross the line to recover the ball. However, the next throw must be made from behind the line.

One trial is usually sufficient on this test. Second trials should be repeated, however, in case of interference of any type or in case the ball gets entirely out of control. It should not be repeated simply because of fumbling. If time and facilities permit administration of two or three trials for all, the higher score should be used.

5. DASH (4 SECONDS)

It is desirable to have a straight course at least 85 to 90 feet long and 4 feet wide. It may be laid out diagonally across the gymnasium if space is too short otherwise. The starting line should be at least 3 feet in front of the wall. The course is marked in one-yard zones beginning at 10 yards from the starting line to about 27 yards from the starting line. The additional distance allows the runner space in which to stop.

Description: Start in any position you wish with the toes behind the starting line. On the signal, Ready, Go! start running as fast as possible and keep going as fast as possible until the whistle blows. You may run as far as you wish after the whistle sounds (at the end

of 4 seconds). The score is the distance you have run between the starting signal and the whistle.

Suggestions: One trial is sufficient unless there is outside interference.

The judge on this should be carefully trained. It is best to use two persons, a timer and a judge. The timer starts the runner and blows the whistle. The judge determines the zone into which the foremost part of the body extends when the whistle blows. One person may assume both responsibilities after training. In this case, the watch is extended forward so the watch and the runner are in a straight line of vision at the end of the 4 seconds.

The judge should attempt to be parallel with the runner at the final signal. With very little experience the judge will learn by looking at the runner during the start and the first few strides whether the finish will be short, around 13 to 15 yards; long, around 23 to 25 yards; or somewhere in between.

If space permits more than one lane, additional lanes should be used providing sufficient judges can be obtained. (See p. 37.)

SCORING THE BATTERY

T-scales for each of the tests and for either battery are shown in Tables XI, XII, XIII. The scores for high school girls may be read from Table XI, those for college women from Table XII. Table XIII is for professional students (majors) in physical education.

The composite on these batteries may be computed in either of two ways. The simplest and quickest is to take the average of the T-scores earned on the three or four tests given. For example, if a student's T-scores are obstacle, 58; basketball throw, 62; broad jump, 60; then the composite score representing the level of motor ability is the average of the three T-scores, or 60.

The second method uses the regression equation derived from the multiple correlation. For the three tests the equation is:

2. basketball throw + 1.4 broad jump — 1. obstacle race

If the four tests are used without the obstacle race the equation reads: .7 basketball throw + 2. dash + 1. passes + .5 broad jump

The actual score may be multiplied by the proper weighting and

TABLE XI

T-Scales for Motor Ability Tests for High School Girls

T-Score	Wall Pass (410) [c]	Basketball throw (ft.) (310) [c]	Broad jump (in.) (287) [c]	4 Sec. Dash (yds.) (398) [c]	Obstacle race (sec.) (374) [c]	Junior High G.M.A.[a] (161) [c]	Junior High G.M.A.[b] (161) [c]	Senior High G.M.A.[a] (169) [c]	Senior High G.M.A.[b] (169) [c]	T-Score
80	16	71								80
79			96-97							79
78								150-151	226-229	78
77	15	68	94-95	27		148-149	205-209			77
76		66			18.5-18.9		201-204			76
75		65					196-200			75
74		64	92-93			146-147	194-195	148-149		74
73	14	63				140-145				73
72		61				134-139	192-193		222-225	72
71		59	90-91	26			190-191	146-147	218-221	71
70		55	88-89		19.0-19.4	132-133		142-145	214-217	70
69	13	54					188-189	140-141	208-213	69
68		52	86-87	25		130-131		138-139	204-207	68
67		51			19.5-19.9		186-187	136-137	198-203	67
66		50				128-129	180-185	134-135	194-197	66
65		49					178-179	132-133	192-193	65
64		48	84-85	24	20.0-20.4	126-127	176-177	130-131	190-191	64
63	12	47				124-125	174-175	128-129	188-189	63
62		46	82-83		20.5-20.9	122-123	172-173		182-187	62
61			80-81				168-171	126-127	178-181	61
60		45		23		120-121	166-167		172-177	60
59		44	78-79		21.0-21.4		162-165	124-125	170-171	59
58	11	43				118-119	160-161	122-123	168-169	58
57		42	76-77		21.5-21.9		158-159		166-167	57
56		41				116-117	156-157	120-121	164-165	56
55		40	74-75	22		114-115	154-155	118-119	162-163	55

54					22.0-22.4		152-153		160-161	54
53		39					150-151	116-117	158-159	53
52	10		72-73			112-113	148-149		154-157	52
51		37			22.5-22.9		146-147	114-115	150-153	51
50		36		21		110-111	144-145		148-149	50
49		35	70-71				142-143	112-113	144-147	49
48			68-69		23.0-23.4	108-109	138-141	110-111	142-143	48
47		34	66-67				136-137		138-141	47
46	9	33			23.5-23.9	106-107	134-135	108-109	136-137	46
45		32	64-65	20		104-105	132-133	106-107	134-135	45
44		31			24.0-24.4		130-131		132-133	44
43			62-63			102-103	126-129	104-105	128-131	43
42		30			24.5-24.9		124-125		126-127	42
41	8	29	60-61	19		100-101	122-123	102-103	124-125	41
40		28					120-121		120-123	40
39			58-59		25.0-25.4	98-99	118-119		118-119	39
38		27	56-57				114-117	100-101	116-117	38
37	7		54-55		25.5-25.9	96-97	112-113		114-115	37
36		26			26.0-26.4	94-95	110-111	98-99	110-113	36
35			52-53	18	26.5-26.9	92-93	108-109		108-109	35
34		25	50-51		27.0-27.4		106-107		106-107	34
33									102-105	33
32		24	47-49		27.5-27.9	90-91	104-105	96-97		32
31	6	23				88-89	102-103		100-101	31
30			44-46		28.0-28.4		100-101		96-99	30
29		22		17	28.5-28.9	86-87	96-99		92-95	29
28					29.0-29.4	84-85	94-95	94-95	90-91	28
27		21			29.5-29.9					27
26			40-43		30.0-30.4	82-83			88-89	26
25	5	20					92-93	92-93	84-87	25
24				16	30.5-31.4					24
23		19	36-39		31.5-32.4	80-81			78-81	23
22				15	32.5-34.9					22
21		16								21
20	4			14	35.0-36.0					20

[a] = .7 basketball throw + 2. dash + 1. passes + .5 broad jump.
[b] = 2. basketball throw + 1.4 broad jump — 1. obstacle race.
[c] = Indicates the number of subjects on which the scale is based.
[d] = Junior High—7th, 8th, and 9th grades.
Senior High—10th, 11th, and 12th grades.
Subjects were from junior and senior high schools in Beaver Dam, Wisconsin; University High School. Iowa City, Iowa; Muscatine, Iowa; and Vinton, Iowa.

TABLE XII

T-Scores for Motor Ability Tests for College Women

T-Score	*Wall Pass (1187)* [c]	*Basketball Throw (ft.) (1162)* [c]	*Broad Jump (in.) (1167)* [c]	*4-Second Dash (yds.) (1173)* [c]	*Obstacle Race (sec.) (1230)* [c]	*G.M.A. (a) (1880)* [c]	*G.M.A. (b) (1228)* [c]	*T-Score*
84	18							84
83		75	84-85	29	17.0-17.9			83
82					18.0-18.9			82
81			82-83				206-207	81
80		65			19.0-19.9	146-147	204-205	80
79	17						202-203	79
78		64	80-81		20.0-20.4	144-145	194-201	78
77		63		28		142-143	192-193	77
76	16	62				140-141	190-191	76
75		60				138-139	188-189	75
74	15	59				136-137	186-187	74
73		58	78-79		20.5-20.9		184-185	73
72	14	57		27		134-135	182-183	72
71		56	76-77			132-133	180-181	71
70		55			21.0-21.4		178-179	70
69	13	54	74-75			130-131	176-177	69
68		53				128-129	170-175	68
67		52		26	21.5-21.9		168-169	67
66		51	72-73			126-127	164-167	66
65		50			22.0-22.4	124-125	162-163	65
64	12	49	70-71	25		122-123	160-161	64
63		47			22.5-22.9		156-159	63
62		46	68-69			120-121	152-155	62
61	11	45			23.0-23.4	118-119	150-151	61
60		44	66-67	24		116-117	148-149	60
59		43			23.5-23.9		146-147	59
58		41	64-65			114-115	142-145	58
57		40		23	24.0-24.4	112-113	140-141	57
56		39					136-139	56
55		38	62-63		24.5-24.9	110-111	134-135	55
54	10	37				108-109	132-133	54
53		36	60-61	22	25.0-25.4		130-131	53
52		35				106-107	126-129	52
51			58-59		25.5-25.9	104-105	124-125	51
50		34					122-123	50

TABLE XII (Continued)

T-SCORES FOR MOTOR ABILITY TESTS FOR COLLEGE WOMEN

T-Score	*Wall Pass (1187)* [c]	*Basketball Throw (ft.) (1162)* [c]	*Broad Jump (in.) (1167)* [c]	*4-Second Dash (yds.) (1173)* [c]	*Obstacle Race (sec.) (1230)* [c]	*G.M.A. (a) (1880)* [c]	*G.M.A. (b) (1228)* [c]	*T-Score*
49		33	56-57	21	26.0-26.4	102-103	120-121	49
48	9	32				100-101	116-119	48
47		31			26.5-26.9		114-115	47
46			54-55	20	27.0-27.4	98-99	110-113	46
45		30					108-109	45
44			52-53		27.5-27.9	96-97	106-107	44
43		29					102-105	43
42	8	28	50-51	19	28.0-28.4	94-95	100-101	42
41		27			28.5-28.9	92-93	98-99	41
40		26	48-49		29.0-29.4		96-97	40
39						90-91	94-95	39
38			46-47	18	29.5-29.9		90-93	38
37		25			30.0-30.4	88-89	88-89	37
36	7		44-45		30.5-30.9	86-87	84-87	36
35		24		17			82-83	35
34		23	42-43		31.0-31.4	84-85	80-81	34
33					31.5-31.9	82-83	78-79	33
32		22	40-41			80-81	76-77	32
31	6	21		16	32.0-32.4		74-75	31
30			38-39		32.5-32.9	78-79	72-73	30
29		20			33.0-33.4		70-71	29
28			36-37	15	33.5-33.9	76-77	66-69	28
27	5	19					64-65	27
26			34-35		34.0-34.4	74-75	62-63	26
25		18	32-33	14	34.5-34.9		60-61	25
24					35.0-35.4	72-73	58-59	24
23	4	17	30-31	13			56-57	23
22		16				70-71	54-55	22
21		15			37.0-37.4		38-39	21
20	3				37.5-37.9	68-69		20
19	2					66-67		19
18		13				64-65		18
17	1		24-25	12				17
16					38.0-38.4			16
15						62-63		15

[a] = .7 basketball throw + 2. dash + 1. passes + .5 broad jump.
[b] = 2. basketball throw + 1.4 broad jump − 1. obstacle race.
[c] = Indicates the number of subjects on which the scale is based. University of Iowa students over a five-year period.

TABLE XIII

T-SCALES FOR MOTOR ABILITY BATTERIES FOR PHYSICAL EDUCATION MAJOR STUDENTS

T-Score	*G.M.A. (1)* [a] *(263)* [c]	*G.M.A. (2)* [b] *(178)* [c]	*T-Score*	*G.M.A. (1)* [a] *(263)* [c]	*G.M.A. (2)* [b] *(178)* [c]
79	166-up		45	116-117	144-147
75	156-160	229-up	44		140-143
73	154-155		43	114-115	
71	152-153	224-228	42	112-113	
70	150-151	220-223	41		136-139
69	148-149		40	110-111	
68		216-219	39	108-109	132-135
66	146-147	212-215	38		128-131
65		208-211	37	106-107	
64	144-145	204-207	36		124-127
63	142-143	200-203	35	104-105	120-123
61	140-141	196-199	34		116-119
60	138-139	192-195	32	102-103	
59	136-137	188-191	31	100-101	
58	134-135		30		112-115
57		184-187	29	98-99	
56	132-133	180-183	27	96-97	
55	130-131	176-179	26	92-95	108-111
54		172-175	25	90-91	96-107
53	128-129		23	86-89	
52	126-127	168-171	21	84-85	92-95
51		164-167			
50	124-125	160-163			
49	122-123	156-159			
47	120-121	152-155			
46	118-119	148-151			

[a] = .7 basketball throw + 2. dash + 1. passes + .5 broad jump.
[b] = 2. basketball throw + 1.4 broad jump − 1. obstacle race.
[c] = Indicates the number of subjects on which the scale is based. University of Iowa students.

these products added. It will be faster, however, to use Table XIV. For example,

basketball throw	35 feet
dash	20 yards
passes	10
broad jump	61 inches

It is quicker to look at Table XIV and find the value of .7 of the basketball throw as 24.5 rather than to multiple it out for each case. Doubling the dash score or taking half of the broad jump can

be done mentally; with little time and effort the total score can be added. Table XIV also provides a multiplication table of 1.4 broad jump.

TABLE XIV

MULTIPLICATION TABLES FOR MOTOR ABILITY TEST BATTERIES

3 Item Battery

2.0 Basketball throw (feet)
+ 1.4 Broad jump (inches)
— 1.0 Obstacle race (seconds)

4 Item Battery

.7 Basketball throw (feet)
+ 2.0 Dash (yards)
+ 1.0 Ball pass (times)
+ .5 Broad jump (inches)

BROAD JUMP X 1.4

Raw score	*x 1.4*	*Raw score*	*x 1.4*
32	44.8	57	79.8
33	46.2	58	81.2
34	47.6	59	82.6
35	49.0	60	84.0
36	50.4	61	85.4
37	51.8	62	86.8
38	53.2	63	88.2
39	54.6	64	89.6
40	56.0	65	91.0
41	57.4	66	92.4
42	58.8	67	93.8
43	60.2	68	95.2
44	61.6	69	96.6
45	63.0	70	98.0
46	64.4	71	99.4
47	65.8	72	100.8
48	67.2	73	102.2
49	68.6	74	103.6
50	70.0	75	105.0
51	71.4	76	106.4
52	72.8	77	107.8
53	74.2	78	109.2
54	75.6	79	110.6
55	77.0	80	112.0
56	78.4	81	113.4

BASKETBALL THROW X .7

Raw score	*x .7*	*Raw score*	*x .7*
20	14.0	45	31.5
21	14.7	46	32.2
22	15.4	47	32.9
23	16.1	48	33.6
24	16.8	49	34.3
25	17.5	50	35.0
26	18.2	51	35.7
27	18.9	52	36.4
28	19.6	53	37.1
29	20.3	54	37.8
30	21.0	55	38.5
31	21.7	56	39.2
32	22.4	57	39.9
33	23.1	58	40.6
34	23.8	59	41.3
35	24.5	60	42.0
36	25.2	61	42.7
37	25.9	62	43.4
38	26.6	63	44.1
39	27.3	64	44.8
40	28.0	65	45.5
41	28.7	66	46.2
42	29.4	67	46.9
43	30.1	68	47.6
44	30.8	69	48.3

It would seem to make very little difference which method is used, that is the average of the T-scores or the weighting of raw scores, since they yield composites which correlate very highly.

EVALUATING THE BATTERY

Let us now analyze these tests and the method of administration which has been outlined. As measures of innate capacity and educability the following points seem significant:

(1) The obstacle race presents skills relatively new as a speed event, yet within the range of experience so that there is no question as to the problem presented.

(2) The obstacle race presents a sequence of movements which is a test of the person's ability to remember directions and to adjust for the next movement while still performing a preceding one.

(3) The obstacle race puts a premium on weight control, balance, total body coordination, and agility.

(4) By avoiding specific instructions on how to throw the balls or make the jump, the tests measure more adequately than would otherwise be possible the performer's knowledge and ability or powers of observation acquired through previous training or experience.

As measures of achievement and general ability to perform, the following points should be noted:

(1) The broad jump is related to leg strength, coordination of arms and legs, and an understanding of the use of effort and balance with respect to one's own body movement.

(2) The basketball throw involves strength, coordination of body and arms, ball handling, and an understanding of the use of effort with respect to some other object.

(3) The wall pass is primarily ball handling, including eye-hand coordination, speed of reaction, and an understanding of the reaction of balls at different angles and speeds.

(4) The dash is considerably more than a pure speed event. Because of its brevity, the start is a very important element. The person who makes a good start and gets up speed quickly covers more distance. The person who gets a slow start does not have time to make up for that slowness. This is a matter of weight control and force as well as reaction time and is very similar to the situations presented in most sports where there are many quick starts.

(5) The parallel lines used for the basketball throw for distance

give an advantage to those who can control the ball sufficiently to deliver it in a straight line.

Since we have defined general motor ability as inclusive of both aptitude and achievement, the two being impractical and undesirable to separate, this battery would seem to be especially suited to measurement of that general ability.

Disadvantages of the tests are not numerous and can be largely overcome by proper administration.

(1) The tests must all be administered individually except the wall pass. They may all be given simultaneously if there is sufficient help, or one each on successive days and the rest of the space used for regular class activity. Most of the mass tests which are sometimes used require one or two class periods to give and are, therefore, no more economical. If there is sufficient assistance the tests may be given as a part of the physical examination though the students usually do not get sufficient warm-up under these circumstances.

(2) Students may learn about the tests and practice on them. This can be prevented by giving them to successive classes without leaving intervening class periods with opportunities for practice.

SUBSTITUTIONS IN THE BATTERY

Other tests may be substituted for these if facilities prohibit, or for any other reason these seem inadvisable. Suggestions for substitutions are as follows:

1. Shuttle Race

Form 1. Parallel lines 15 feet apart. Score is the number of times the performer can cross between the lines in 15 seconds.

Form 2. Parallel lines 25 feet apart marked in 5-foot zones. (See Figure 29.) Work in partners.

Description: Start at line X, run to line Y, change direction and return to X. Repeat this as many times as possible in 30 seconds. Record the number of times your partner runs each length of the course and note the letter of the area the runner reaches at the end of the 30 seconds. (Example, 10-B)

Suggestions: It is important that the runner wait until the whistle blows, and that lines X and Y be touched each time, and that the scorer record the area that the runner is in at the second whistle.

The shuttle race may be substituted for the obstacle race or the dash.

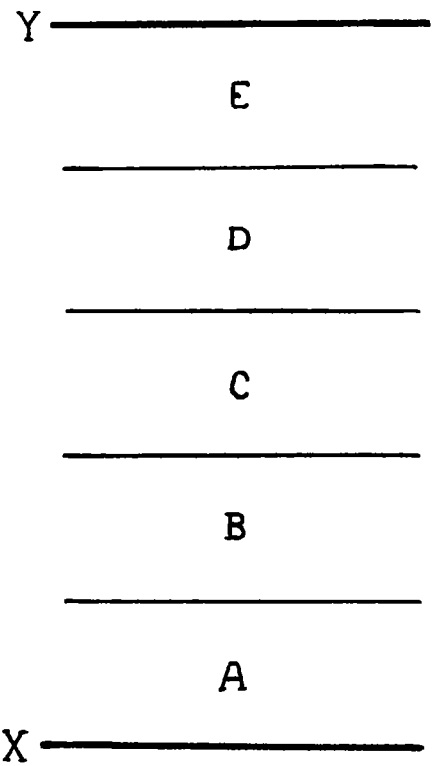


Figure 29. Floor Markings for Shuttle Race

X — starting line
X-Y — lines on which to reverse direction
A, B, C, D, E — zones in the shuttle area

2. JUMP AND REACH

Stand facing the wall, toes touching with both hands raised overhead. Reach evenly with both hands and mark the height of the reach. Then turn sideward to the wall, jump and reach with one hand touching as far up the wall as possible. The score is the difference between standing and jumping reach. If the wall is such that water will show on it and not leave the wall disfigured, the

easiest method is simply to dip the fingers in water when starting the test and then measure promptly. If this is not feasible, short pieces of chalk must be used.

With college women the original reach may be measured more accurately if they stand with backs to the wall. The assistant marks the reach.

This test may be substituted for the broad jump if mats or pit are not available for the jumping.

3. SAND BAG THROW

Use sand bags 4 inches square weighing 1 pound each, and with a string tied tightly around the middle of each to prevent its being thrown so that it will "sail" through the air. The administration of the test is identical with the basketball throw for distance in all other respects.

This test may be substituted for the basketball throw for distance where balls are lacking or the space is short.

USING THE MOTOR ABILITY SCORES

The motor ability score is used to section classes, to arrange squads or teams, or to determine the level of achievement expected in future work, and the amount of individual help which will be needed to achieve satisfactory results. For that reason the tests should be given at the beginning of certain phases of the student's physical education experience. This does not mean that it needs to be given every year. It is probably best to give this battery at the beginning of the junior high school years, at the beginning of the senior high school course, and when entering college. The results will be used then for a two or three year period. Improvement doubtless occurs during this length interval but the relative standing of students will not change appreciably, unless some students practice specifically on these test items while others do not.

One specific example of the value of general ability tests lies in the economy of time for later testing. Placement by the general test can be used in connection with every activity, and therefore permits a very much shortened battery at the beginning of each

activity season for classification in that activity. The general battery gives a fairly good estimate of the level of ability to be expected of a person who is just starting to learn a new activity. If later achievement in that activity shows markedly superior status from that predicted it can almost always be attributed to extra practice, to exceptional motivation, or unusual interest and effort. Likewise, distinctly lower status than that predicted can usually be found to be due to lack of effort and practice.

Another use of the motor ability measures is in selecting the groups at both ends of the scale for special consideration. For example, the lowest 15 to 25 per cent in ability too frequently are simply spotted as dubs and left to shift for themselves as best they can in the class. The result is discouragement, dislike for the activity, and eventually lack of cooperation, not to mention the fact that they usually remain poor in skills. If those individuals can be given special help, preferable in classes by themselves for a time, most of them profit considerably.[11] They may improve on specific skills by special help, they may learn to analyze skills more carefully, to compensate for some of their shortcomings in capacity or aptitude, and above all they have the opportunity to practise skills in a sympathetic group where they do not feel unnecessarily self-conscious.

The upper 15 to 25 per cent in ability can also profit by special consideration. When all levels of ability are taught in the same class, by the same procedure, at the same rate, the most capable ones are not challenged by the material presented and often discouraged because they work and play constantly with others of very poor skills. Those with high general ability may be selected for leaders classes, given additional skills or projects on which to work, placed in advanced classes, or placed in special classes with others of like ability. The latter procedure is very frequently followed in college courses and these groups make phenomenal progress in most activities.

In short, the students of low ability can be taught to work hard, to achieve modest results and to work objectively on shortcomings, and to understand that if certain skills are to be passed or standards met, that they must spend more time and effort than others. For example, beginning swimmers who are afraid of the water would

be perfectly willing to concede that more effort is necessary to learn to swim than for those who are not afraid. Likewise, the persons who have little apparent aptitude for complex motor skills are just as aware of their difficulty as the frightened swimmer. If they are motivated to achieve some degree of skill, they will recognize the need for practice and be willing to exert the necessary effort.

In a similar manner, the student of high ability can be inspired by the variety of skills which may be acquired or the high level of skill which may be achieved. In both cases the students are aided in setting up their own goals, their effort is directed always toward improvement, as well as in surpassing some one else.

BIBLIOGRAPHY

1. Alden, Florence, Horton, Margery, and Caldwell, Grace: "A Motor Ability Test for University Women for the Classification of Entering Students into Homogeneous Units," *Research Quarterly*, 3, March 1932, p. 85
2. Hatlestad, Lucille: "Motor Educability Tests for Women College Students," *Research Quarterly*, 13, January 1942, p. 10
3. Johnson, Granville: "Physical Skill Tests for Sectioning Classes into Homogeneous Units," *Research Quarterly*, 3, March 1932, p. 128
4. Kistler, J. W.: "The Homogeneous Grouping of Junior and Senior High School Boys for Physical Education Class Activities," *Research Quarterly*, 8, December 1937, p. 11
5. McCloy, C. H.: "The Measurement of General Motor Capacity and General Motor Ability," *Research Quarterly*, 5, Supplement March 1934, p. 46
6. ——: "Recent Studies in the Sargent Jump," *Research Quarterly*, 3, May 1932, p. 235
7. ——: "An Analytical Study of the Stunt Type Test as a Measure of Motor Educability," *Research Quarterly*, 8, October 1937, p. 46
8. Metheney, Eleanor: "Studies of the Johnson Test as a Test of Motor Educability," *Research Quarterly*, 9, December 1938, p. 105
9. Niehaus, Marian: *A Study of Tests for Dividing Junior High School Girls into Homogenous Groups for Physical Education.* Unpublished M.A. thesis, University of Iowa, 1935
10. Powell, Elizabeth, and Howe, Eugene C.: "Motor Ability Tests for High School Girls," *Research Quarterly*, 10, December 1939, p. 81
11. Salit, Elizabeth Powell: *Development of Fundamental Sport Skills in Freshman College Women of Low Motor Ability, Research Quarterly*, 15, December 1944, p. 330
12. Scott, M. Gladys: "The Assessment of Motor Ability of College Women," *Research Quarterly*, 10, October 1939, p. 63
13. ——: "Motor Ability Tests for College Women," *Research Quarterly*, 14, December 1943, p. 402

7.

Achievement Ratings and Progressions

There are some activities which do not lend themselves to objective testing. In spite of that fact it may be desirable to know the relative status of members of a class, or to motivate effort by noting rate of improvement. This can be made possible by subjective, systematic ratings of the performance of each person. These subjective ratings may serve as a substitute for tests. Dancing or diving are examples where subjective ratings only might be made. In other cases the ratings may serve as a supplement to the tests. An example of this combination might be found in tennis. Tennis tests may show accuracy of placement but usually fail to discriminate between strokes varying in speed, flight, bounce, and other factors influencing the ease with which an opponent might return the ball. A test of accuracy combined with a rating of form would be much more valuable than the test alone.

Other activities are of a self-testing type, such as swimming or bowling. If there are many separate units or skills constituting the activity, then some kind of progression scale serves to indicate desirable sequence of learning and to afford discrimination in the rate at which achievement proceeds. Ratings and achievement progressions have certain elements in common. However, they will be discussed separately, scales and charts provided in those activities where they will be most useful.

RATINGS

Certain preliminary preparations must be made before ratings can proceed. These are inter-related.

(1) The scale or range must be determined. If the activity is not easily judged for differences in efficiency or ease, then a "pass"

or "fail" category may be sufficient. Usually better results will be obtained if at least three categories are used; i.e., the "fail" is retained, the "pass" is subdivided into fair and good. The words, which are descriptive, may be used or numerical values may be substituted, as fail = 0; fair = 1; good = 2. Or, if no failures are to be anticipated the scale may read, poor = 1; fair = 2; good = 3. Then by adding the rating given on two or more items or skills, a composite score is obtained, with the highest number indicating greatest ability. The scale or range may be increased to five points but it is seldom advisable to set up more than five categories because discrimination between like cases becomes too difficult and too time consuming.

(2) Each point on the scale must be clearly defined. This definition must be in terms of the particular activity and describe a level of ability which will be found in the class. For example, in rating a swimming stroke:

3—*good,* position good, coordination and timing good, uses drive well, and keeps resistance at a minimum; easy, relaxed stroke.

2—*fair,* position reasonably good, coordination correct but lacks proper rhythm and force.

1—*poor,* stroke is recognizable but position is poor or inconsistent, too much effort with little result, lacks ease and relaxation, can swim width of pool.

0—unable to swim across the pool maintaining proper floating position and using stroke throughout.

If more categories are added then the analysis of coordination and timing is carried further with certain types of errors holding the swimmer down on the scale in contrast to other errors which might be considered less serious.

(3) The opportunity for rating must be planned. Ratings should not be made from memory of what the performer does; but rather the performer should be seen in action. In case of swimming it might mean watching the swimmers in small groups or individually while they try to swim a specified distance. In case of tennis it might mean watching the player in a game on the court, or stroking against a backboard. Rating should then be

done as objectively as possible, judging the performance and forgetting the personality and previous impressions which that person has made.

(4) Score sheets should be planned for greatest ease in rating and totalling ratings. Plans should also be made for the student to see his ratings, especially if opportunity is to be given for further practice on those skills.

(5) If names of the performers are not thoroughly familiar to the teacher or the person making the rating, a number should be pinned on each performer in order that the rating may be done accurately and promptly.

RATING OF DIVING

Diving is judged subjectively whether in competitive meets, in trying for swimming and diving awards, or in regular class instruction and rating. This subjective rating is made reasonably accurate by setting up the important elements of the dive and a scale which determines the proper classification for a performance with various combinations of these elements. The following discussion illustrates this procedure for diving in general. In some cases of special dives additional specific points may be added.

ELEMENTS OF GOOD FORM IN DIVING

1. The position of readiness is erect and well balanced.
2. The approach is legal (at least three steps in case of a running dive).
3. The approach is smooth, easy, and in good posture.
4. The hurdle and take-off are timed with the board.
5. The body is straight, with effective arm and leg action at the take-off.
6. The height of the flight is sufficient to permit the necessary movements or position.
7. Body position during flight conforms to specification of the dive (i.e., tuck, pike, lay-out, twist, or number of revolutions).

8. All body movements during flight are smooth, easy, and limited to those which are essential.
9. The entry into the water is with body straight, close to the line of the take-off.

For regular class use a scale of five points is recommended. It should be arranged as follows:

5—*excellent,* dive meets all specifications for good form, no apparent errors which call for further coaching.

4—*good,* dive gives a general impression of good form, minor variations exist which would improve the dive if corrected.

3—*average,* dive meets the basic specifications but lacks smoothness and ease, or lacks control in some one respect which affects the dive as a whole. (For example, too much forward lean on the take-off decreases height and moves the point of entry farther from the take-off.)

2—*fair,* dive is inadequate and full of errors, but has some indication of control, or merit in some aspect.

1—*poor,* dive is recognizable but fails to meet the standard in practically every element involved.

Competitive diving rules usually require that dives be judged on a ten point scale. Such a scale merely differentiates within each of these five categories and gives corresponding points. The above scale may be broken down as follows:

5 = 8½, 9, 9½, or 10
4 = 6½, 7, 7½, or 8
3 = 4½, 5, 5½, or 6
2 = 2½, 3, 3½, or 4
1 = ½, 1, 1½, or 2

However, in judging competitive diving the dive is automatically scored as zero if it is not the particular variation of the dive which is entered or announced. Competitive diving further varies the points awarded according to the difficulty of the dive. This has nothing to do with the judging. The dive is usually given the average of the ratings of three judges each working independently. Let us assume three different dives each of which received an

average rating of 8 from the three judges. The rules may specify that one dive is classed as 1 in difficulty, another is more difficult and classed as 2, the third is still more difficult and classed as 3. The number of points won by these three dives is the judges' rating, 8, multiplied by the difficulty value 1, 2, or 3. Therefore, the dives win points 8, 16, and 24 respectively. Usually this procedure is not followed in class work where students are apt to be working on dives similar in difficulty.

The judges will have the most complete view of the dive if one stands at the side approximately in line with the end of the board, a second on the opposite side a little beyond the end of the board, and the third stands back near the rear end of the board.

It is understood in competitive swimming that diving judges must be highly qualified, that the same judges rate all divers, and that they each follow the exact specification and scale prescribed for the dives. Under these circumstances, the ratings by the various judges will be quite consistent and a satisfactory means of evaluating the performance. Exactly the same procedure is necessary as a part of a swimming test. (See Chapter 9 for a discussion of ratings by experts.)

RATING OF POSTURE

Evaluation and rating of posture has never been thoroughly satisfactory. One method that is used is to ask the person to stand in either a good or habitual position and have one or more persons rate that position. This procedure is made slightly more objective by the use of plumb lines or vertical lines in the background and by indication of certain landmarks. A very valid objection to this procedure is that one is moving much more of the time than standing, and that ability to stand well does not always insure ability to move well.

Another alternative for the same type of posture rating is to take a posture picture or silhouette and make a rating of the picture. This has certain advantages, namely, the record is permanent and better comparison of similar figures and more uniformity of rating will be obtained, (2) the picture can be taken at successive times

and direct comparisons made, (3) the picture provides the student with an opportunity to see how he looks. The disadvantages are that it is a stationary pose, that there is still no objective criterion or measure to be employed, and it is expensive and impractical to use in some cases.

The main objection to these two methods is partially overcome by asking the person to perform specific movements and the rating is made of that performance. This may be done in succession in a class or may be done informally by watching the person in daily activity doing such things as walking, sitting, going up or down stairs, or carrying books or other loads. The latter has the advantage of seeing the person naturally in activities and emphasizing the need for good posture at all times; but has the disadvantage of being time-consuming for the rater and difficult to see all persons in a sufficient number of situations.

One solution to most of these drawbacks is suggested in the plan for class testing discussed below. This is a plan for the organization of the class and selection of activities in which posture is to be rated. Whether standing position only is to be rated, or simple and complex movements are to be included, the outline of elements in posture must be available for the rater. Therefore, let us consider such an outline first.

ELEMENTS IN GOOD STANDING POSTURE

1. Back and Head
    1. The three curves of the spine should be moderate.
    2. The head should appear well balanced on top of the spine with the line of vision and the chin horizontal.
    3. The trunk should appear easily erect without being stiff.
    4. The outline of the sternum and ribs should be more or less straight in front, with a long vertical axis, rather than sunken, collapsed, and with a concave axis.
    5. The spine should appear straight when viewed from the rear.

2. ABDOMEN AND PELVIS

1. The line of the abdomen should be straight or very mildly convex. (There is an exception to this in the small child up to about six years of age. The abdominal line is usually convex but the following point still applies.)
2. The abdominal wall should be mildly firm, not relaxed and sagging, and not stretched or containing excessive fat deposits.
3. The pelvis should be held squarely beneath the trunk, not with the lower back and abdominal wall projecting forward at a pronounced angle.*

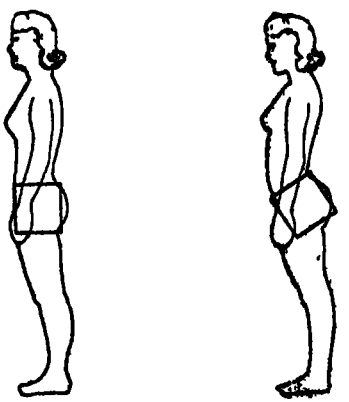


Figure 30. Illustration of Pelvic Alignment

4. The lower line of the abdominal wall should be behind the lower end of the sternum.

3. SHOULDER GIRDLE AND ARMS

1. The shoulder blades should not deviate appreciably from the contour of the spine and thorax. When

* The pelvis may be considered as a box, even though it is distinctly irregular in shape. (See Figure 30.) That box should be kept squarely aligned as though resting on a level surface and not tipped up on edge. Tilting the pelvis invariably causes the buttocks to project backward more prominently just as it causes increased concavity of the lumbar spine. However, care must be taken to judge pelvic alignment by the position of the pelvis itself and not the spine and abdomen above it. The contour of the buttocks is not an accurate means of judging alignment since there is great variation in the muscular development and hence of the outline of the hip. Figure 30 illustrates the same girl with the pelvis in the two positions.

viewed from the side as in a posture picture they should not markedly exaggerate the convex line of the upper back.

2. The shoulder girdle should be carried far enough back that the arms hang easily at the sides.
3. The shoulder girdle should be retracted without giving the appearance of stiffness and without thrusting the ribs and sternum forward unduly.
4. The shoulders should be low, not shrugged or tense.

4. Feet and Legs

1. The feet should be parallel.
2. The inner line of the feet should be straight rather than convex (not pronated).
3. The heel cord in the rear should be straight, not turned in at ankle level.
4. The upper surface of the feet should appear straight or convex, not sunken and spread just back of the toes.
5. The knees should be straight without rigidity, not flexed.

5. General Alignment and Weight Control

1. The following landmarks should be situated one above the other when the person is viewed from the side: lobe of the ear, point of the shoulder, hip joint, rear of the patella.
2. The line through these four points should be vertical and extend downward through the feet midway of the base (heel to ball of foot) which brings it a little in front of the ankle joint.
3. General appearance is of relaxation and control rather than rigidity.

When the rating is made during movement most of the same points still apply and the following are added.

WALKING

1. Contact for each step should be made first with the heel.
2. Push-off at each step should come from the toes, principally the great toe. Failure to get a drive from the toe gives an appearance of rocking over the foot and then simply lifting it without noticeable ankle and foot action.
3. Leg action should be free, without tenseness and without conspicuous swaying of the hips in either a lateral or vertical direction.
4. Arms and shoulders should be relaxed but controlled within a relatively small arc.
5. Balance should be maintained over the base without stiffness, or flexion at the hips or hyper-extension in the lower back.
6. There should be no jar to the body at the moment of heel contact.

RUNNING

1. Contact should be made first with the ball of the foot.
2. The push-off from the toes should be strong.
3. Vertical bobbing should be eliminated by knee action.
4. Excessive forward reach by the legs should be eliminated by bringing the leg downward and backward for the step rather than stretching it out for a long stride.
5. Body lean should be greater with increased speed, but should result from the body inclining forward as a whole, not by trunk or hip flexion.

MISCELLANEOUS ACTIVITIES

1. The body, including weights to be lifted or carried, should always be in optimum position of balance, resulting in conservation of muscular effort and strain.
2. Excessive range of movements and superfluous movements are always undesirable.
3. Relaxation should be as complete as is possible for good body mechanics in the task involved.

A very practical and useful diagnostic test might be organized as in the following chart which would take care of a squad of 10 girls. That number can be scored very conveniently at one time.[6]

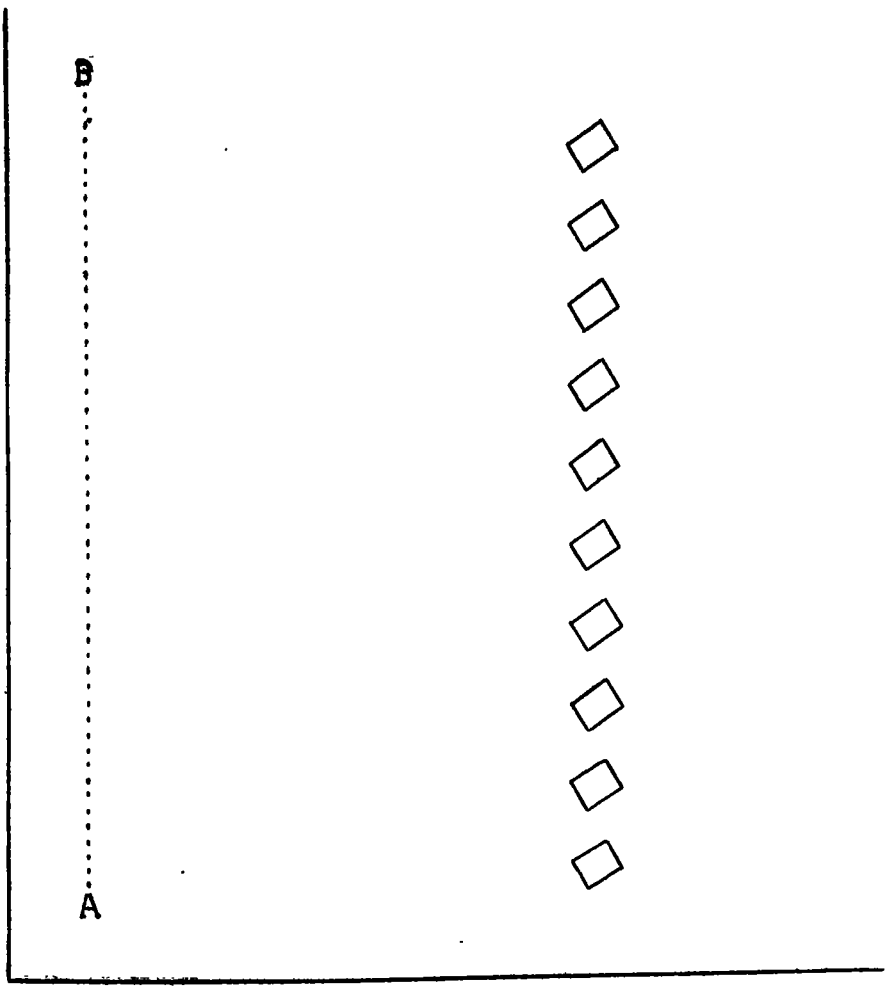


Figure 31. Illustration of Arrangement for Administering Posture Test

A-B Area in which examiner moves to observe each student. The space between A-B and chairs is for students to perform activities for the rating

A three point scale could be used; 3—good, 2—fair, 1—poor. Standing and walking posture would doubtless be included. Use of the feet in walking is most easily judged separately from the general rating of walking. The additional items to be included in

the test would probably be chosen from activities such as running, stair climbing, sitting, stooping, reaching overhead, carrying a load, pushing or pulling.

In preparing for the test place ten chairs in line, one in front of the other, with a little space between. There should be some open floor space beside them. Names of squad members are entered on the score sheet; when the test is given students are seated in the same order as in the list and they remain in the same order throughout the test. Adequate rating can be given only if the students are dressed in swimming suits or tight fitting suits, and are barefooted.

The examiner stands to one side of the row of chairs.*

1. Each girl in turn walks a few steps toward the examiner, turns and walks away again. This gives opportunity for judging the foot alignment and pronation.

2. Each girl in turn walks a few steps forward (with side to the examiner). This gives opportunity to examine for heel contact, weight transfer and toe drive.

3. Each girl stands in line while the examiner moves down the line rating standing posture.

4. The girls walk two or three at a time back and forth beside the line of chairs. During this the examiner rates the walking posture. Having more than one walk at a time helps to avoid self-consciousness and unnatural gait.

5. Girls sit in the chair in a natural sitting position for rating. Each rises and then sits again to be judged on balance and movements.

6. Movement on stairs should be both up and down. The test may be given on real stairs, preferably wide ones for a better view and to accommodate two girls at a time; or it may be given on stairs constructed for this use in the gymnasium.

* The test and chart represent modifications of a test developed by the Staff of the Department of Physical Education for Women, University of Iowa. They are similar to forms published by that department, to one appearing in Posture and Body Mechanics by Loraine Frost, University of Iowa Extension Bulletin; and to another in Tests and Measurements in Health and Physical Education by C. H. McCloy, F. S. Crofts and Company. Quoted by permission of each of the above publishers.

Name	*Foot Mechanics*			*Standing*	*Walking*	*Running*	*Sitting*		*Stairs*			Total Score
	No pronation	Feet Parallel	Heel to toe contact	Alignment & relaxation	Alignment & ease	Alignment and coördination	Alignment seated	Rising, Sitting	Up Down	*	*	

* To be filled in with choice of activities.

Figure 32. Sample of Posture Test Score Sheet

Other items in the test can be set up in a similar manner preferably with some properties or setting to make the movements seem natural.

RATINGS IN DANCE

Ratings represent about the only approach to measurement in dance, and are used for evaluating ability in the various types of dance. They can be made on dance steps, such as the waltz or schottische, or on the performance of an entire dance, as might frequently be the case in square dances, for example. It is better to have just a small group to be watched at one time; to alleviate embarrassment, it may be better to have several persons dancing but only a few being judged.

The rating scale on a dance might read as follows:

5—*excellent,* is skillful in steps involved; knows positions and floor pattern, and sequence of steps; shows a feeling of assurance and of enjoyment, and expresses the spirit of the dance.

4—*good,* executes the steps correctly in form, rhythm and sequence, but is lacking something in ease, naturalness, or expression.

3—*average,* performs the dance correctly with only minor errors which she is able to correct herself or by cues from other dancers, apparently knows it well enough to enjoy it.

2—*fair,* can execute the basic steps involved and can perform the dance reasonably well by the lead of partner or other dancers; rhythmically she makes an occasional error and is unable to readjust readily. She may or may not enjoy it depending upon whether she is disturbed by her own errors.

1—*poor,* performs dance steps poorly; she is almost entirely dependent upon others for cues for sequence; rhythmically she is inaccurate and apparently unaware of it.

RATING OF SOFTBALL BATTING FORM

The construction of objective skill tests in batting has been made more difficult by the lack of a mechanical device which would deliver the ball in the same manner to all batters. The ability of the pitcher is known to affect the ability of the batter. Since batting comprises the major part of the offensive action in a game, it is important that it be measured, and in as nearly an objective manner as possible. Having the same pitcher deliver the ball to all batters being tested has been tried; usually the number of balls that must be delivered is too great for the pitcher's endurance. Batting is a skill in which the element of chance plays a large part, and therefore the number of pitches to each batter has to be quite great. Such tests are time consuming and take considerable space. The use of a rating form during actual playing time seems to be the best substitute at present.

The rating form can be mimeographed, with one form cumulative in nature, for each player. For use in instruction, these forms should be ready for use the early part of the season. They can be handed to the player after each turn at bat. The same form can be used in a few tournament games at the end of the season, with the tally marks being recorded in a different color on successive ratings so that improvement can be noted. The ratings will be fairly reliable if each batter is rated in three games at the start of the season and two or three at the end. Student leaders can be taught to do the rating; it is important that they understand the fundamentals of good batting form.

The rater should stand behind the plate umpire and slightly toward first base, in foul territory. The use of several raters is preferable. (See Ratings, p. 232.)

The detailed form presented here does not include the result of each batter's turn at bat. Such a record can be obtained from the score book or seasonal batting averages, and should not unduly influence the rating of form. If time is limited the checking of errors can be omitted. Their inclusion is mainly to facilitate using the blank as a teaching device.

Date	*Rater's Initials*
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

INSTRUCTIONS:

1. Fill in date and your initials on each form.
2. Rate the player each time she bats. Place a tally mark in the space which precedes the best description of player's form (good, fair, poor) in each of the six categories.
3. Indicate your observation of errors on the right hand half of the page, again with a tally mark.
4. Write in any additional errors, and add comments below.

Player's name:_____

Captain's name:_____

Class Hr., Days:_____

RATING	ERRORS
1. GRIP	_____Hands too far apart. _____Wrong hand on top.
_____Good	
_____Fair	_____Hands too far from end of bat.
_____Poor	
2. PRELIMINARY STANCE	In relation to plate, stands:
	_____too near it. _____Rear foot closer to plate than forward foot.
_____Good	_____too far from it.
	_____too far forward toward pitcher.
_____Fair	_____too far backward toward catcher.
	_____bat resting on shoulders. _____Shoulders not horizontal.
_____Poor	
3. STRIDE OR FOOTWORK	_____Fails to step forward.
	_____Fails to transfer weight.
_____Good	_____Lifts back foot from ground.
_____Fair	_____
_____Poor	_____
4. PIVOT OR BODY TWIST	_____Fails to twist body.
	_____Fails to "wind up."
_____Good	_____Has less than 90° of pivot.
_____Fair	_____
_____Poor	_____
5. ARM MOVEMENT OR SWING	_____Arms are held too close to body.
	_____Rear elbow held too far up.
	_____Bat not held approximately parallel to ground.
_____Good	_____Batter does not use enough wrist motion.
_____Fair	_____Wrists are not uncocked forcefully enough.
_____Poor	
6. GENERAL	_____Batter's movements are jerky.
(Eyes on ball, judgment of pitches, etc.)	_____Batter tries too hard; "presses."
	_____Fails to look at exact center of ball.
_____Good	_____Poor judgment of pitches.
_____Fair	_____Batter appears to lack confidence.
_____Poor	_____Poor selection of bat.

ADDITIONAL COMMENTS:

_____

_____

Figure 33. Sample Rating Sheet for Softball Batting Form

BASKETBALL RATINGS

It is sometimes desirable to rate basketball players in order to get estimates of skill or improvement in some particular phase of the game. This may be done by watching the players in the game, and judging them on a three or five point scale similar to other scales illustrated in this chapter. Many of the basketball tests in use involve basket shooting because they give objective scores. Other skills do not lend themselves so readily to objective testing.

In the women's game, it has been found by observation and some experimentation that skill in use of the bounce is very closely related to general skill in the game. Such a rating is particularly good to help in classifying players for beginning or advanced groups, for classes, or placement on teams. This can be rated most readily by having two or three players working together on one half of a basketball court. They are asked to pass the ball from one to another, to cover the floor as though working through a defense. Each is also asked to use a bounce to herself as frequently as possible. The bounce may be judged as follows:

3—*good,* keeps ball under control at all times; covers distance, keeps ball low, avoids all fumbling and traveling.

2—*fair,* uses the bounce to advantage; but does not cover as much distance as might be desirable, or bounces at a height which would be intercepted easily, or occasionally travels.

1—*poor,* gets no advantage from the bounce, and would probably lose the ball by interception or traveling in the game, bounces straight down to the floor, or travels consistently.

General footwork may be judged in the same way in small groups of two or three. The additional points to be considered would be ability to start, run, stop, or change direction quickly by means of a reverse turn or pivot; to judge speed and position with respect to space on the floor and action of teammates' play; to avoid traveling when in possession of the ball, but to be able to keep the play

moving. The main justification for this type of rating is (1) that it is very difficult to evaluate in actual tests, and (2) that it approaches the game situation more adequately than many tests. The player's rating in this case is much more apt to be affected by the ability of the partner being tested than is true in rating the bounce. For that reason it may be desirable sometimes to change pairings for this rating.

RATINGS IN TENNIS

Accuracy of placement of the serve can be tested objectively; likewise, ability to keep a ball in play. (See skill test, p. 100.) Form must be rated subjectively. This can be done while the players are playing regular games, using the scale below as a guide. It is presumed that the students are divided into two ability groupings for instructional purposes and that the rater will have had contact with the group over a period of time.

1. For Beginners

   5—*good,* executes all strokes in good form. May have played before this term, has learned rapidly, perhaps worked outside of class or watched others play. Profits by all suggestions.

   4—*above average,* plays the game sufficiently well to avoid being conspicuous on the courts for poor playing. Has shown definite improvement and is anxious to learn.

   3—*average,* shows fair but somewhat inconsistent form; knows the essentials of the game, scoring, etc.

   2—*near dub,* can stroke in fair form but is careless; has improved some during the term but has little knowledge of the game.

   1—*dub,* has poor strokes and has made little progress. Makes little effort to improve.

2. For Intermediates

   5—*expert,* has good form and plays consistently well, is fast, knows the game and uses excellent strategy,

and can play either singles or doubles reasonably well. This does not imply that the player is an experienced tournament player.

4—*near expert,* usually has good form and plays smart tennis; perhaps plays singles better than doubles, or vice versa; is somewhat erractic but knows the game.

3—*average,* usually plays a good game, tries out new tactics, and is analytical concerning own game. The greatest need is practice.

2—*fair,* is weak on some techniques but has ability to im-improve, understands weaknesses, and has knowledge of game procedure.

1—*poor,* has not overcome poor habits of technique, has little knowledge of the game or analytical ability.

The above scale is general in nature and can be used by various raters even though they disagree about certain points on form. Another method of rating which is more specific but is limited to the ability to execute strokes is described below. Its chief advantage is that it is more economical of time then the first method, which measures the ability of the player to use his skills and knowledge in a game situation.

Divide the players into small groups for rating the serve and supply each with several balls. Line them up along the baselines, all on the same side of the net. (For indoor rating, place the baseline at official distance from the net or backboard allowing about five to ten feet between players.) As many as ten can be judged at the same time. Their names should be listed on the score sheet in the same order as the arrangement of players. The number of repetitions necessary before the raters make final judgment will depend largely on the experience of the judges; but the entire process, rating the three basic strokes, should not take more than sixty activity minutes for a class of forty players.

For the forehand and backhand, other players are needed for putting the ball in play. The class can be divided into three groups, with the *throwers* stationed near the net, the *strokers* behind the baseline on the opposite side of the net from the throwers, and

the *ball chasers* behind the throwers. When this rating is conducted indoors, care must be taken to protect the ball throwers from possible injury. The ball chasers can be eliminated and the number of strokers may need to be reduced. The ball throwers can be asked to kneel on the floor behind the net. Since the ability of the ball throwers affects the ratings, a rotation plan should be used, to insure that no one player be handicapped throughout the test by having to work with a poor thrower.

When the game is subdivided, with each stroke being rated separately, a chart similar to that in Figure 34 can be used. It can be extended to include the volley and lob. Rate on a three or five point scale.

Name	*Forehand*	*Backhand*	*Serve*		

Figure 34. Sample Chart for Tennis Ratings

ACHIEVEMENT PROGRESSIONS

Some activities build naturally from one skill or ability to another. Usually such classes are organized so that different members may work more or less at their own rate. Items are then checked off as they are accomplished and the student knows that he is then ready to advance to other or more difficult items. This is true in both swimming and tumbling.

Proper use of such a chart requires that the student know immediately the results of his effort. Also, opportunity should be given to study the chart, to know what to start on next, to know what should be accomplished eventually, and to set own goal or rate of work. Such a procedure can be a powerful motivator.

CLASS: SQUAD: NAMES:	*Jelly-fish float	Face float and recovery	Back float and recovery	Rhythmic breathing	Elem. back stroke—60 ft.	*Flutter kick—20 ft.	Seal dive from side	Surface dive—shallow water	*Kneeling dive from side	*Side stroke—20 ft.	Vertical float—deep water	Treading water—30 sec.	*Elem. crawl—20 ft.	5 minute swim	Standing dive—deep water	Side stroke—60 ft.	Deep water shuttle	Treading—2 min.	Crawl—60 ft.	15 min. swim		

* Optional items.

Figure 35. Sample Achievement Chart for Elementary Swimming

Jelly-fish float. Swimmer assumes tuck position and holds while partner counts 10, recovers standing position without help.
Face float. Swimmer assumes straight prone position and holds while partner counts 20, recovers standing position without help.
Rhythmic breathing. Standing with trunk bent forward, face in water, arms moving as for crawl stroke (or windmill fashion). Exhale under water, turn head and breathe in; face under water and exhale. Repeat rhythmically at least 10 times. (This should be taught with the same timing between arm action and breathing as used later in the crawl.)
Elementary back stroke. Continue regular stroke full length of pool. (Swimmers are not sent full length of pool until the instructor passes the stroke for deep water swimming.)
Flutter kick. Swimmer assumes face float with push-off from the side, uses rhythmical and symmetrical flutter kick until she reaches the other side (most pools for instruction are 20 or 25 feet wide).
Five minute swim. Swimmer uses any stroke. The start is in the water and swimming or floating continues without touching sides or bottom. One timer on the side can time all, starting and stopping the interval with a whistle.
Deep water shuttle. Swimmer jumps in feet first, levels off and swims any stroke 15 feet, stands, submerges till the head is entirely under water, comes up and levels off, swims back 15 feet with another stroke, does a swimming turn (i.e., turns while stroking in a horizontal position), swims by means of a third stroke 15 feet, stands and treads water for at least 3 kicks, levels off and swims back to the starting point. (Note that this requires the use of all three strokes learned but they may be used in any order.)
Fifteen minute swim. Dive in head first, swim or tread continuously for the first 7½ minutes; float, swim, or tread for the last 7½ minutes.

A progression chart naturally reflects the teaching plans; i.e., the level of skill expected from the group, and the steps by which the skills are to be achieved.

Charts of the type illustrated here can usually be used by members of the class on each other, or by squad leaders. Simply specify the standard which constitutes success. If more discrimination is needed and adequately trained assistants (squad leaders) are available, it is no more time consuming to rate as fair and good, i.e., 1 or 2.

Examples from different activities follow which illustrate the points already discussed.

SWIMMING

The achievement chart for elementary swimming is for a series of eighteen to twenty-four lessons for a beginning class being taught swimming as a safety and recreational skill. Some members of the class may not finish the last three or four items. However, items of that difficulty are essential if those who learn more rapidly are to be stimulated to real effort. If members of the class start with some ability, then obviously each should finish those which are starred, since these represent merely progressions into something else.

The standards for passing each item might be similar to those fairly detailed chart and if time is limited omit items such as the elementary chart and start on an intermediate one. This is a below the chart and are simple enough to be judged by class members.

STUNTS AND TUMBLING

The tumbling stunts are divided into groups according to the principal skills involved. The progression chart includes stunts of increasing difficulty within each group. The various groups are usually carried along simultaneously, with students making more rapid progress in one group of stunts than in others. This chart is simply illustrative of method and is not meant to indicate the events to be taught.

	flexibility						balance							agility							strength						co-ordination					
CLASS:	*Wicket walk*	*Crab walk*	*Worm walk*	*Through the stick*	*Jack knife*	*Stork stand*	*Dwarf walk*	*Head stand*	*Hand stand*	*Forearm stand*	*Chest balance*	*Standing mount*	*Bear dance*	*Top*	*Forward roll*	*Backward roll*	*Wind the clock*	*Jump the stick*	*Walrus walk*	*Seal slap*	*Single squat*	*Horse and rider*	*Dead man's lift*	*Elbow dip*	*Knee stand to standing*	*Heel jump*	*Fish flop*	*Cart wheel*	*Hand spring*	*Snap up*	
SQUAD:																															
NAMES:																															

Figure 36. Sample Achievement Chart for Stunts and Tumbling

The stunts may be checked off when achieved, or ability may be evaluated by a two or three point rating scale. The first procedure is probably best if the students check each other, or if the squad leaders do the checking. The rating scale is preferable if the instructor is doing the rating. It is possible to weight the stunt for difficulty as in the case of the diving. This is good if conducting competition between individuals or squads, but is not essential for regular class work.

BOWLING

The bowling score in itself is indicative of relative ability of players. However, it does not analyze strong and weak points in one's game. To be of greatest value in teaching, a record should be kept of some of the set-ups which occur frequently in the game. The number to be included in the chart will depend upon the time allotment. This is an excellent practice device.

If practice is exclusively on regular games the only set-up which they face with any consistency is the full ten pin arrangement. The percentage of strikes or spares can be kept on total frames rolled. However, a chart of this type gives the score on special set-ups for the second ball or spare attempt.

An individual score sheet with a continuous record is preferable. That is the form indicated here. This makes an excellent teaching device and gives a diagnostic record for both teacher and student.

ARCHERY

Archery is a self-testing activity and it is scored in such a way that every arrow shot is rated for its relative accuracy. Scoring also permits a partial analysis of performance, i.e., whether the shooting is consistent or variable, whether one distance is more difficult than a shorter or longer one. The officially recognized rounds provide exactly the same kind of a score as that obtained from tests in other sports. Other rounds or series of scores may be used in the same way. The scores on a given round may be used as a measure of performance, or scores on the same round may be compared at intervals to measure improvement. For the college student

BOWLING CHART

Name _____

One pin set-ups

4	5	6	7	7	10	10					

Two pin set-ups

1-3	1-2	3-5	4-7	5-7	5-8	5-9	5-10	6-10	7-9	8-10	

Three pin set-ups

1-2-4	1-3-5	1-3-6	1-2-9	1-5-8	2-4-5	3-5-6	4-5-7	4-7-10	5-8-9		

Four pin set-ups

1-2-4-7	1-2-4-9	1-2-7-9	1-3-5-6	1-3-6-9							

Five pin set-ups

1-2-3-4-5	1-2-4-7-10	1-3-6-8-10

Set-ups left when bowling games

Date											

Instructions:
If the whole set-up is cleared, score as a spare ⊠.
If some pins are left standing, record the pin numbers of those still up. If the whole set-up is missed, record as O.
Use the blank spaces for new set-ups on which you need more practice. Take at least five successive balls at the same set-up. Repeat on later days until you are at least 80% accurate.

Example

6-10	
⊠	0 ⊠ 6 10

Both pins down on the first and third trials, both pins missed on the second trial. Pin 6 left on the fourth, and 10 left on the fifth. This is only 40% accurate. Practice the set-up again.

Instructions:
When bowling regular games, record here the numbers of the pins in each set-up left for your second ball. This will indicate the set-ups on which you need most practice.

Figure 37. Sample Chart for Bowling Practice

Hyde presents standards of performance on the first and last Columbia Round, and therefore suggests the amount of improvement which may be expected from students.[5]

BIBLIOGRAPHY

1. Alexander, John: "A Motivating Individual Record," *Journal of Health and Physical Education,* 10, December 1939, p. 582
2. Bennett, LeVerne Means: "A Test of Diving for Use in Beginning Classes," *Research Quarterly,* 13, March 1942, p. 109
3. Buhl, Olga Anderson, and Morrill, Warren P.: "The Measurement of Postures," *Research Quarterly,* 12, October 1941, p. 518
4. Crabtree, Helen Kitner: "A Test for Riding," *Journal of Health and Physical Education,* 14, October 1943, p. 419
5. Hyde, Edith I.: "An Achievement Scale in Archery," *Research Quarterly,* 8, May 1937, p. 109
6. Frost, Loraine: "Posture and Body Mechanics," *University of Iowa Extension Bulletin, No. 479,* March 1940
7. Judd, Mary: *A Study of the Distribution of Weight on the Foot Walking with and Without Shoes.* Unpublished M.A. thesis, University of Iowa, 1943
8. Streit, W. R.: "A Stunt Meet for Elementary School Boys," *Journal of Health and Physical Education,* 10, December 1939, p. 584
9. Wickens, J. Stuart, and Kiphuth, Oscar W.: "Common Postural Defects of College Freshmen," *Research Quarterly,* 13, March 1942, p. 102

8.

Construction of Knowledge Examinations

Written examinations have not been as widely used in physical education as have some other types of tests, although the teachers who have made use of them have found them to be very helpful. Certainly all of us agree that one of our objectives is the acquiring of knowledge and most of us agree that a grade in a course should reflect the degree to which the student has progressed in all of the objectives. We are not content to base grades on attendance records alone; we should not base them only on improvement in skill. Practically all of our courses have a knowledge content, although the amount varies.

AVAILABILITY

Very few knowledge tests in physical education have been published. None of the specially organized test building agencies, such as the Cooperative Test Service and the various state and regional testing bureaus, has prepared tests in physical education. When the demand for such tests is sufficient, doubtless the test building agencies will undertake to meet it. Until such tests are available, teachers must construct their own and use as guides the few good ones that have been published.

USES

As pointed out in Chapter 1, written examinations serve many purposes other than as a partial basis for assigning marks. The teacher can use them to discover what the student knows at the outset, and therefore upon what level to begin instruction. For example, the teacher of a rhythm course may want to know how much the students know about music.

Another very important use is as a teaching device. Short quizzes can be used to stress certain points, or to aid students in summarizing knowledge of units within a course. Some teachers make good use of them in teaching activities where the rules are somewhat complex, as for example, basketball, field hockey, and softball.

One of the newer and most important uses of written tests is that made by some of our teacher training departments who are attempting to determine the needs of their entering students. They recognize that the backgrounds of these entering students vary greatly and that it is not wise to require the same courses of all. For example, if the knowledge and skill of an entering student is beyond that of the "average" student who has completed the basic elementary course in any activity, then that student is permitted some form of election. The written test is used in connection with skill tests, and other information that the advisor has, to determine the requirements. It is obvious that knowledge tests that are used for this purpose must be very carefully constructed. Such tests usually require more time and effort than the average teacher can give, as the various questions need to be analyzed for their efficiency and the tests rebuilt on the basis of earlier results.

SOURCES OF INFORMATION

For the teacher who must prepare his own examinations, suggestions are given here. Much has been written on the construction of knowledge tests in such subject matter fields as history, English, and mathematics. Considerable of that which has been written is applicable to tests in our field, and assistance can be gained by studying some of the better texts.[4, 9] The discussion here will be confined to the field of physical education.

DISTRIBUTION OF CONTENT

The first step is to determine the use to be made of the examination. If you are going to give the test as a final examination in a course or use it for classification, then care must be

[4, 9] See bibliography at end of this chapter for these and other references.

taken to make it be comprehensive. The course outline should be consulted, and if it is brief, the test constructor will need to elaborate it, listing all of the important concepts. In general, the larger number of questions should be devoted to those things that are considered to be of prime importance. Avoid overweighting the examination with questions covering just one phase of instruction, as for example, rules of an activity. Following the course outline rather than the text or rule book will help to avoid this all too common failing.

The teacher who plans to use a ready-made test should consult the content distribution to see if it coincides with the emphases he has made in his course. An example of an attempt to distribute the number of questions according to content is that found in a soccer knowledge test, prepared for professional students.

	Number of Questions	%
Analysis of individual techniques (how to do in good form)	3	5
Analysis of game situations and use of skills	8	13
General knowledge (history, selection and care of equipment, safety precautions, differences between field ball, speedball, and soccer)	3	5
How to avoid fouling	3	5
Placement of passes, throw-ins, kicks for goal	2	4
Tactics and areas of play	22	36
Rules essential to intelligent play	15	25
Terminology	4	6
	60	

CHOICE OF TYPE OF ITEM

After deciding the proportions, the next step is selecting the type of item that fits the content best. The types most useful for physical education will be discussed and illustrated.

MULTIPLE CHOICE

1. Forms

There are several forms of multiple choice exercises or items in use.

Form A: A direct question followed by a number of responses, only one of which is correct and all others definitely incorrect.

Example, from a field hockey examination:

What is the umpire's decision when the ball is sent over the endline, last touched by a member of the attacking team from within the striking circle?

(1) Long corner.
(2) Penalty corner.
(3) Penalty bully.
(4) Twenty-five yard line bully.
(5) Free hit.

Here there is but one correct answer, number 4, and all of the others are definitely incorrect.

Form B: A direct question followed by a number of responses, all or some of which are acceptable in various degrees but one of which is definitely better than any other. This is known as the "best-answer" type.

Example, from a volley ball test for girls:

What is the best use that can be made of the set-up?

(1) To place the ball in position for a spike.
(2) To pass the ball to a teammate.
(3) To remove the spin from the ball so that it can be played more accurately.
(4) To encourage good teamwork.

Here the first answer is definitely better than the others, although all are acceptable to some degree.

Form C: An incomplete statement with several possible completions provided, one of which is to be selected.

Example, From a tennis test:

In the parallel system of court coverage in doubles, the server and her partner should

(1) Take positions at the net imme-

diately following the delivery of the first serve.

(2) Remain behind the baseline until they are able to force the opponents into a disadvantageous position.

(3) Go to the net as the ball is being returned to their side of the net.

(4) Assume positions about halfway between the baseline and the service line, waiting for an opportunity to go to the net.

Here the answer will depend on what the instructor has taught the class, 1 or 2 being considered best, 3 and 4 being definitely incorrect.

Form D: An identification type of question, with a list or key of abbreviations for the choice of answers placed at the top, and then a series of questions, with blank spaces provided for answers in the left hand column.

Example, from a basketball test:

Directions: Select the appropriate answer and place the symbol or abbreviation in the blank space immediately preceding the question. P, personal foul; T, technical foul; TC, technical foul charged to the position of captain; V, violation; L, legal play.

—— 15. A player along the free throw lane steps into the lane as soon as the ball leaves the hands of the free thrower.

—— 16. A player catches a ball while running and takes three steps before stopping.

This type of question has been widely used in the Women's National Officials Rating Committee examinations. The description of the situation should be brief, yet give all the necessary information. All the symbols should be used at least once, and they should be repeated at the top of each page to avoid errors or waste of time in answering the questions. This is particularly important in an activity where the symbols are somewhat confusing, as in volley

ball, where "S" may be used for "side out" and "SO" for "serve over."

There are other forms with minor modifications, mostly typographical in type. The direct question form, either A or B, is better than the incomplete statement (Form C) in that the student knows from the outset what problem is being presented and is saved the time of rereading the stem in connection with the responses. This is particularly true when the sentence is long. Form B is preferable to Form A in that it tends to test for deeper knowledge. In Form B, the student must read all of the responses and then decide which one is best.

There is no particular advantage in having a fixed number of responses. It may make the test look neater but it tends to cause the test constructor to throw in some responses that may not function, that are merely superfluous. If only three plausible responses can be contrived, then use only three. Three is the minimum and there is seldom any advantage in having more than five.

2. Uses

The multiple choice type of test item seems to fit the content of most of the subject matter in physical education and is preferred for the following reasons:

1. They can be adapted to test for any depth of understanding.

2. They can be made completely objective in scoring and are easily adapted to answer sheets. (See answer sheet, p. 46.)

3. It is possible to detect readily any non-functional material in the responses, thus facilitating the revision of the question for later use. (A non-functional response is one which is never selected.)

4. They test the student's ability to eliminate incorrect responses as well as to select the correct response directly.

5. They do not require correction for guessing.

6. They seem to have fewer disadvantages than the other commonly used forms: alternate response (true-false, yes-no, etc.), matching forms, and recall (completion, analogy, definitions, etc.).

Use multiple choice when many alternatives are possible, but some basis exists for superiority of one. Use it when students can be expected to differentiate between closely related points.

3. Rules for Construction

1. Use a short, simple, direct question form for the stem.

2. Avoid choices which are not plausible or which are too obvious.

3. If the directions call for selecting the one correct answer, avoid having more than one correct response. (It is preferable to state in the directions that the best response is to be sought; that there may be more than one but that one will be superior.)

4. Avoid answering one question by another.

5. Avoid unintentional clues, such as placement of correct or best response consistently in a certain place in the series. Another clue is word matching between the stem and the response. Making the correct response consistently longer or shorter than the incorrect should be avoided. An example of a grammatical clue is the use of a singular expression in the stem and plural ones in all but the correct response. Another grammatical clue is the use of an incomplete statement in the stem ending in "a" or "an." (In a survey of teacher made examinations, more grammatical clues were found in those questions employing an incomplete stem than in any others.)

6. Avoid use of textbook language, if you wish to test for ability to use information, for understanding, and not just memorization. It is considered legitimate to use familiar or stereotyped phrasing in an incorrect response occasionally, to deliberately mislead the shallow thinker.

4. Negative Approach

Some times the material may adapt itself to a negative approach, although the negative approach is to be avoided if possible. It may be used when the test constructor can find many more correct responses than incorrect ones. At times it is almost impossible to decide what the best answer would be to a positive form of the question. If the negative form is to be used for a number of questions, it is advisable to segregate them and label them as negative. A substitute for this is to italicize or underline the negative portion of the stem so that the student is sure not to miss seeing it.

An example of a negative question from a soccer examination:

The following statements concern trapping the ball with the sole of the foot. Which one is *false?*

(1) Keep your eyes on the ball rather than on an opponent.
(2) Use this trap for balls that are rolling slowly on the ground.
(3) Put your weight on to the foot that is over the ball, clamping the ball between the foot and the ground.
(4) Place the middle of the sole on the ball, lowering the heel slightly.
(5) Get your body in line with the oncoming ball, so that you face it squarely as it approaches.

Here 3 is a false statement and, therefore, the correct response.

Another example of a negative question is taken from a track and field test.

Which of the following does *not* disqualify a throw in the javelin event?

(1) Stepping on the throwing line before the javelin leaves the hand.
(2) Stepping over the throwing line before the javelin leaves the hand.
(3) Javelin failing to stick in the ground.
(4) Two hands used on the javelin during the approach.

The correct answer is 3.

Still another example of a negative approach, demonstrating the use of the word "least" in the stem, is taken from a body mechanics knowledge test.

Which of the following exercises would you expect to be of *least* value in minimizing painful menstruation?

(1) Airplane exercise.
(2) Bicycle exercise.
(3) Double knee circling exercise.
(4) Prone lying, head and shoulder raising exercise.
(5) Knee-chest exercise.

The answer is 4.

5. SCORING

The number of right responses is the score. This can be quickly obtained when using a superimposed key by counting the errors and subtracting that sum from the total number of items.

ALTERNATE RESPONSE FORMS

In constructing an examination, the teacher will be confronted occasionally with content which seems to have but two possible responses. Included under the general heading of Alternate Response Forms are the true-false, the yes-no, and the multiple response. We will consider each briefly.

1. FORMS

In the true-false, a statement is made and the student indicates whether it is true or false. Usually the provision for the answer is made on the test form itself, with the letters T F preceding the number of the question. The student encircles the T if he considers the statement to be entirely true; the F, if he considers the statement to be partially or wholly false. Sometimes a blank space is provided in front of the question, in which the student places the abbreviation "T" or "F." This, of course, is not as satisfactory as the previously mentioned method, due to illegibility of handwriting. The abbreviation of + for true and — for false has been tried also, but has the same difficulty of not always being legible. If the examination is of such length that it extends over several pages, it will be more economical to use answer forms. The one illustrated on page 46 can be adapted for either multiple choice or alternate response questions, and can be scored with a punched key. If the statement is considered true, the student places an X in the first row of brackets; if false, in the second row. In this way, the same answer sheet could be used for an examination including both multiple choice and alternate response questions. The directions may read, for example, that the first twenty-five questions are of the multiple choice type and questions twenty-six through seventy-five inclusive are of the true-false type.

In the multiple response form of item, the statement is followed by three or more responses. The directions call for selecting

all of the answers which are true. There may be no correct answers; there may be any number of correct answers. An example of an item in which all of the answers are true follows:

Softball. A runner on third base may score on

(a) A wild pitch.
(b) A passed ball.
(c) A foul tip.
(d) A throw back from catcher to pitcher.
(e) A fly ball caught, providing that third base is held until the ball is caught.

A better form of the stem is the direct question. In the question above the stem should read, On what play may a runner on third legally score? The question form is again illustrated below. In this case there are only three correct answers (a, d, e):

Softball. Which of the following players may wear gloves but not mitts?

(a) Pitcher.
(b) Catcher.
(c) First baseman.
(d) Outfielder.
(e) Shortstop.

2. Uses

The alternate response type of item can be used advantageously when you must cover large amounts of information and must economize on the pages of typing, or when little time is to be spent in constructing an examination and its use will be limited to motivation or instructional purposes. Another time when it can be used is when the information is merely factual and testing is for the ability to memorize, and there is no desire to test for the ability to make applications. Alternate response questions can be good or poor.

3. Rules for Construction

1. Make the statements or questions brief and direct.
2. Avoid ambiguities.
3. Avoid textbook wording.
4. Have an approximately equal number of each alternative, with no regular pattern to responses.

4. SCORING

The scoring recommended for true-false, yes-no, and plus-minus types is usually "rights minus wrongs." This is more quickly computed by taking the total number of items minus twice the errors. If you have directed the students *not* to guess, then they should be penalized for errors. Thus, the scoring would be the total number of items minus twice the number of the errors, minus the number of omissions. In the multiple response type. an omission is obviously an error and should be scored accordingly.

RECALL

1. FORMS

Another grouping of forms is the Recall, in which a short response is expected associated with the question raised. One form uses a stimulus word or phrase, with a blank space provided for the student's answer (a single word or phrase).

Example: Measurement of width of hips: *caliper.*

This type of form varies in difficulty from one extreme to the other. If the student supplies an equally good word but not the one the examiner had in mind, it may be marked wrong, thus making the examination tend to be too difficult and more of a guessing contest than a knowledge test. To prevent this from occurring the test constructor may be so definite in the part he supplies that the answer is obvious.

The analogy form is illustrated below:

I.Q.: Intelligence = ———— : Motor Ability

(ANSWER: M.Q.)

It has a very limited use.

The sentence completion, where the student is asked to supply the missing word, is generally considered very poor. It tends to test for the trivial or to be a test of intelligence or vocabulary rather than a test of the content of the course. Its use should be avoided.

A better type is the short essay or sentence, which calls for a single sentence reply or for listing.

Example, from a basketball test:

What are the advantages of shifting zone defense over player-to-player defense? List four.

1.
2.
3.
4.

The difficulty here is in the scoring. The player may list four but they may not be the four which are most important. Or he may combine two under one number, leaving a space blank, or perhaps filling it incorrectly.

2. USE

It is a better form for instructional purposes than for testing purposes. Recall questions are generally considered to be more difficult than recognition questions. Their main use is in identification of terms, or for purely factual information. They could well be used for quizzes in anatomy or kinesiology, in asking for some such information as insertions and origins of muscles.

3. RULES FOR CONSTRUCTION

1. Be sure that there is only one correct answer.
2. Objectivity depends on brevity of the answer.
3. If spaces are provided, they should be of uniform length, or they will serve as a clue. They should be long enough to take care of the longest reply.
4. Space for answers should be provided in or near a margin, to facilitate scoring.

4. SCORING

They should be scored according to the number of correct responses.

MATCHING

1. FORMS

Still another type of question is the matching exercise. This is sometimes classified with multiple choice. There are two common

forms: two columns of single words; or one column of words or names, with one of phrases or explanations.

2. USES

Matching forms may well be used for the "who, when, and where" type of information. It is obviously weak in measuring interpretative abilities.

3. RULES FOR CONSTRUCTION

1. The second column should contain the responses and should always have more items than the first column, to prevent answering the difficult ones on the basis of elimination alone. The items in this column should be numbered.

2. Blank spaces, for recording the number of the matching item should be placed in front of the items in the left hand column.

3. Clues, such as grammatical form, proper names or capitalization, should be avoided.

4. The list should have homogeneous content.

5. The directions should state whether items in column two may be used more than once.

6. The instructions should be specific on the basis on which connections are to be made.

7. The response column should be arranged in sequence, alphabetically or numerically.

4. SCORING

The score is the number of correct responses. If several choices are possible, correction for guessing must be made. (Total items minus twice the errors, minus one for each omission.)

USE OF MISCELLANEOUS DEVICES

DIAGRAMS

Diagrams should be used when questions involve spatial relations, or wherever they can make the situation more clear. Sometimes they actually save space. The examination should be arranged so that all the questions making use of a certain diagram are placed on the same page with it; or the diagrams may be placed on a

separate sheet of paper, each labeled with the question numbers.

The use of diagrams in connection with questions involving the flight of an object is illustrated with questions from a golf examination.

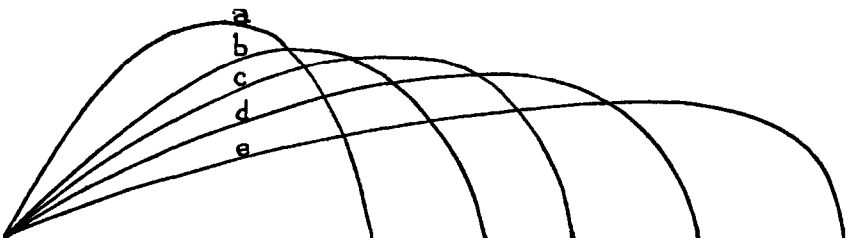


Figure 38. Diagram of Flight of Ball with Iron Clubs

1. Of the flights shown in the diagram (Figure 38) which most closely approximates that of the midiron?
   (1) a.
   (2) b.
   (3) c.
   (4) d.
   (5) e.
2. To make a stroke with a flight similar to *a* (Figure 38), what should you do?
   (1) Stand farther from the ball than for a stroke made with a wooden club.
   (2) Check the follow through in the direction you want the ball to go.
   (3) Take a stance with the ball nearer the rear foot than the front one.
   (4) Keep the face of the club closed.
   (5) Aim just above the center of the ball.

NOTE: Answers: 1, 5; 2, 3.

Sometimes one diagram can be used not only to save words but to save the time necessary for mental imagery. It also has the advantage of providing the player with a more natural situation. An example of such a diagram is taken from a basketball test for girls:

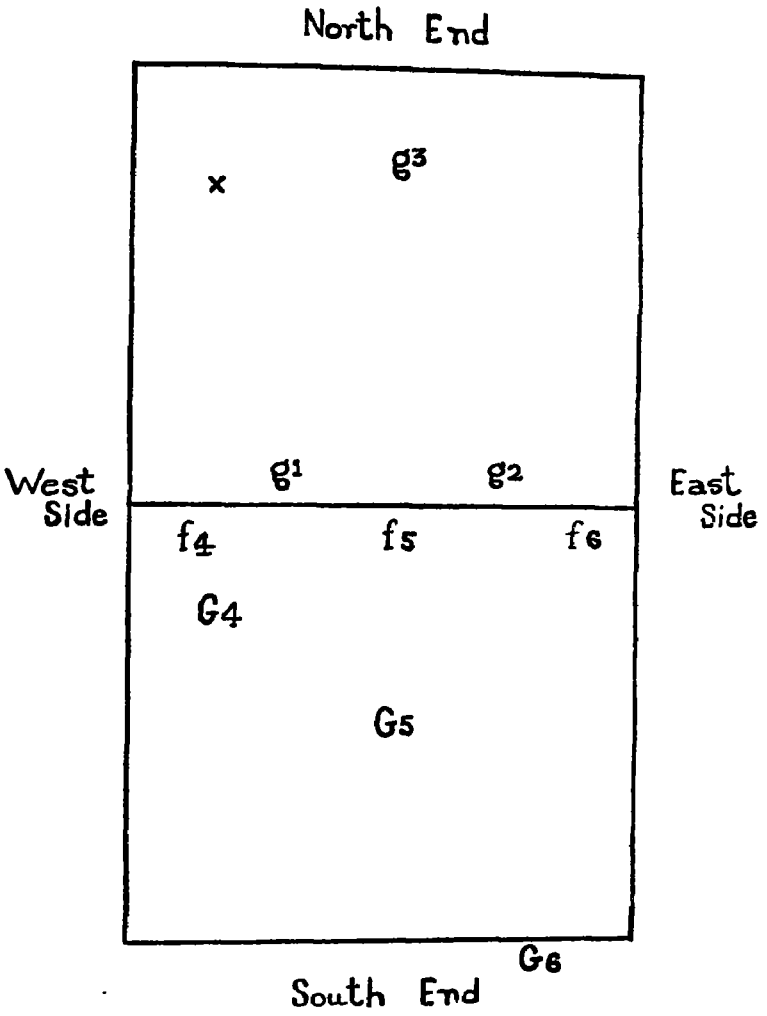


Figure 39. Basketball Diagram

27. (See Figure 39.) G6 has the ball out-of-bounds at the end line. The "small letter" team is employing a shifting zone defense. If the ball is passed from G6 to G_5 to G_4, what should f_5 do?
    (1) Move toward G_5.
    (2) Remain where she is.
    (3) Move toward f_4.
    (4) Move closer to center of division line.

separate sheet of paper, each labeled with the question numbers.

The use of diagrams in connection with questions involving the flight of an object is illustrated with questions from a golf examination.

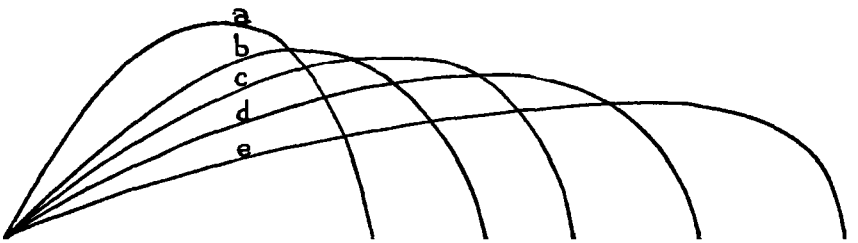


Figure 38. Diagram of Flight of Ball with Iron Clubs

1. Of the flights shown in the diagram (Figure 38) which most closely approximates that of the midiron?
   (1) a.
   (2) b.
   (3) c.
   (4) d.
   (5) e.
2. To make a stroke with a flight similar to *a* (Figure 38), what should you do?
   (1) Stand farther from the ball than for a stroke made with a wooden club.
   (2) Check the follow through in the direction you want the ball to go.
   (3) Take a stance with the ball nearer the rear foot than the front one.
   (4) Keep the face of the club closed.
   (5) Aim just above the center of the ball.

NOTE: Answers: 1, 5; 2, 3.

Sometimes one diagram can be used not only to save words but to save the time necessary for mental imagery. It also has the advantage of providing the player with a more natural situation. An example of such a diagram is taken from a basketball test for girls:

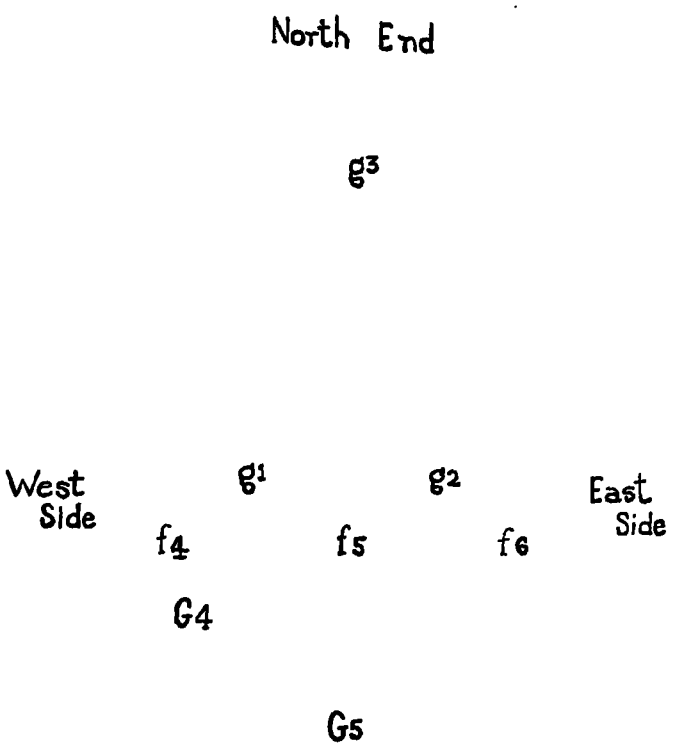


Figure 39. Basketball Diagram

27. (See Figure 39.) G6 has the ball out-of-bounds at the end line. The "small letter" team is employing a shifting zone defense. If the ball is passed from G6 to G_5 to G_4, what should f_5 do?
    (1) Move toward G_5.
    (2) Remain where she is.
    (3) Move toward f_4.
    (4) Move closer to center of division line.

28. (See Figure 39.) If the ball is passed above the reach of g1 to a forward at the spot marked X, what should g2 do?
    (1) Move toward the forward at spot X.
    (2) Remain where she is.
    (3) Move toward the northeast corner.

Answers, depending somewhat on style of defense taught: 27, 3; 28, 3.

SUBSTITUTES FOR DIAGRAMS

Sometimes a tabulation similar to those below can be incorporated in the responses themselves to save lengthy descriptions.

Examples, from bowling:

22. What is the score at the end of the third frame in the following game?

Frame	*1st ball*	*2d ball*
1	6	4
2	5	2
3	8	1

    (1) 26.
    (2) 31.
    (3) 33.
    (4) Correct answer not listed.
    (5) Incomplete.

23. What is the score at the end of the third frame in this game?

Frame	*1st ball*	*2d ball*
1	6	4
2	1	9
3	8	2

    (1) 30.
    (2) 31.
    (3) 39.
    (4) Correct answer not listed.
    (5) Incomplete.

24. What is the score at the end of the third frame in this game?

Frame	*1st ball*	*2d ball*
1	10	0
2	7	3
3	2	7

(1) 29.
(2) 38.
(3) 41.
(4) Correct answer not listed.
(5) Incomplete.

25. At the end of the seventh frame the score was 100. What is the score at the end of the tenth frame?

Frame	*1st ball*	*2d ball*
8	9	1
9	10	0
10	8	0

(1) 128.
(2) 138.
(3) 146.
(4) Correct answer not listed.
(5) Incomplete.

The correct answers: 22, 2; 23, 5; 24, 3; 25, 3.

The same sort of thing can be done in rhythmical form and analysis tests.

Example: 28. What is the time signature for the waltz?

(1) 2/4.
(2) 3/4.
(3) 6/8.
(4) 4/4.
(5) 5/4.

29. In which rhythmical pattern is the time syncopated?

(1)
(2)
(3)
(4)

Correct answers are: 28, 2; 29, 3.

CHECK LIST FOR EVALUATING ITEMS

After the items have been prepared, the use of the following check list * will prove helpful in evaluating them.

(1) Exactly what is this item intended to measure?

(2) Is the intended purpose of the item acceptable? Is it important that the item be included; does it test for something significant?

(3) Is there any ambiguity in the item? Will the student recognize the purpose of the item? Can it be made more clear? Are there any qualifying phrases that might start the student to thinking along an irrelevant line?

(4) Does the item contain any unintentional clues to the correct response?

(5) Will authorities agree on the correct response? Are the responses which are intended to be wrong really less acceptable than the correct one?

(6) In multiple choice items are any of the wrong responses likely to appear more plausible than the correct answers to the best of the students to be tested? Is the item too difficult for the best students in the group?

(7) Is the item phrased as economically as possible? Is it straight forward, direct?

(8) Is the form of the item as well adapted as any to its intended purpose? Would a diagram help?

(9) Would the rote learner have any undue advantage in responding to the item? Has "textbook" language been avoided?

Another method of evaluating questions before trying them on a group is to have some other person go over them. Re-reading the questions yourself after the lapse of a few days is also a good procedure.

For help in the technical problems involved in setting up the examination and administering it, see Chapter 3, pages 44 to 47.

* Check list adapted from class notes: Improvement of the Written Examination, taught by Dr. E. F. Lindquist, University of Iowa.

As a guide to proper administration, use this check list:

(10) Is the provision for the student's response as economical of his time as possible?

(11) Are the directions to the student as simple and understandable as they can be?

(12) Does the provision for the student's response provide for convenient, accurate, and economical scoring?

(13) Can the typographical arrangement of the items be improved?

(14) Is the spread of estimated difficulty of items adapted to the spread of ability in the group to be tested? (See Chapter 9, p. 230.)

(15) Are the time limits adequate?

(16) Are the questions placed in the test in an order progressing from easy to difficult, as estimated? Will the slow student be prevented from spending an undue amount of time on items that are too difficult for him?

If a test is to be a valid test, that is, if it measures what it purports to measure, the following criteria should be considered:

(17) Are any important objectives or outcomes of instruction seriously neglected in the test as a whole?

(18) Is the emphasis on functional value and not on content objectives?

(19) Is there any undue testing of isolated detail or unimportant items of information, such as terminology or definitions, for their own sake?

(20) Are test situations suggestive of the life situations in which the student may make actual use of what he has learned?

(21) Would this test be less satisfactory if used as an "open book" test? Are there a sufficient number of questions which require drawing of inferences and making of applications?

Obviously, these are rigid criteria and it is doubtful if any test ever meets all of them; but they are goals toward which the conscientious may strive. Some may have to be sacrificed somewhat for the sake of others, for example, number ten. The answer sheet requires more of the student's time; it has been estimated that the students can answer about ten percent fewer items when using an answer sheet than when writing directly on the test forms. This percentage is reduced when students become accustomed to using

are used; otherwise scoring is too difficult. Arrange the questions in order of estimated difficulty, ranking from easy to hard. The copy given the typist should be readable and in as nearly as possible the exact form that you wish it. Be sure to:

1. Provide a space for the student's name and other essential data, either on answer sheets, or on test forms if answer sheets are not to be used.
2. Indicate amount of space to be left for diagrams.
3. Ask that the question be arranged so that the diagrams can appear on the same page with the questions that refer to them.
4. Ask that no questions be split; all of the items should appear on the same page with the stem.
5. If abbreviations or symbols are to be used ask that they be repeated at the top of each page.
6. Ask to proofread the stencils before the examination is reproduced.

ADMINISTERING THE EXAMINATION

Suggestions for administering the examination and scoring the papers are included in Chapter 3. Conversion of raw scores into letter grades is covered in Chapter 9.

HOW TO CHECK ON THE EFFECTIVENESS OF EXAMINATIONS

Perhaps a few statements should be made that will aid teachers who wish to revise and improve their tests. It is assumed that the teacher who takes considerable time and care in preparing an examination will want to use it again and will, therefore, not permit the students to retain their examination papers. We will not deal with the evaluation of questions by statistical methods here (see Chapter 9) but will list a few criteria for subjectively determining a general estimate of the worth of the examination in its entirety:

1. Did it provide a wide range of scores, with no undue massing of scores at any one point along the scale?

2. Does the order or rank of scores coincide roughly with your previous estimate of the abilities of individuals within the group?

3. Was the examination sufficiently comprehensive? Did it cover all the important phases of instruction?

4. Was it the right length? Would the students have had time to answer more questions?

5. Were the questions clear? Did the discussion of papers reveal any ambiguities?

6. Did the examination indicate areas of content which need no further instruction? Did it reveal inadequate understanding of certain phases of subject matter?

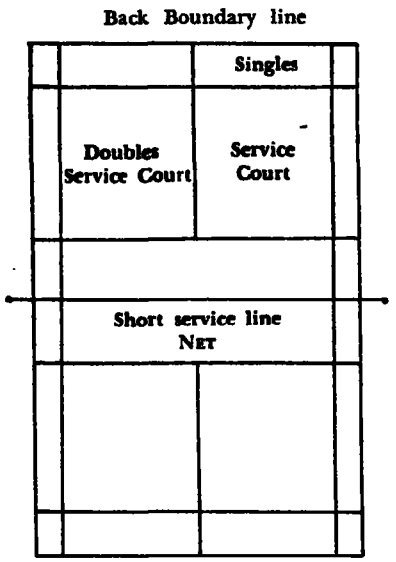
REVISING EXAMINATIONS

In revising examinations, the results previously obtained need to be carefully studied. Sometimes the difficulty of questions needs to be adjusted to the anticipated ability of the group. If you wish the total scores to differentiate between members of a certain section of the class, then quite a few questions with difficulty corresponding to their ability should be included. The following question, from a tennis examination administered to a group of eighty-four physical education major students, proved easy:

> What strokes, in addition to the serve, forehand drive, and backhand drive, are most important for the beginner to master?
>
> (1) Volley, chop.
> (2) Lob, volley.
> (3) Half volley, chop.
> (4) Half volley, volley.
> (5) Lob, chop.

The answer, 2, was selected by seventy-five of the eighty-four, or eighty-nine percent. This gives the question a difficulty rating of only eleven. Of the nine who failed to select the right answer, three selected response 1, two selected response 4, three selected response 5, and one omitted the question. The question was retained for further use because it discriminated well between the good and poor students, but response 3, which was not selected by anyone

Court Areas

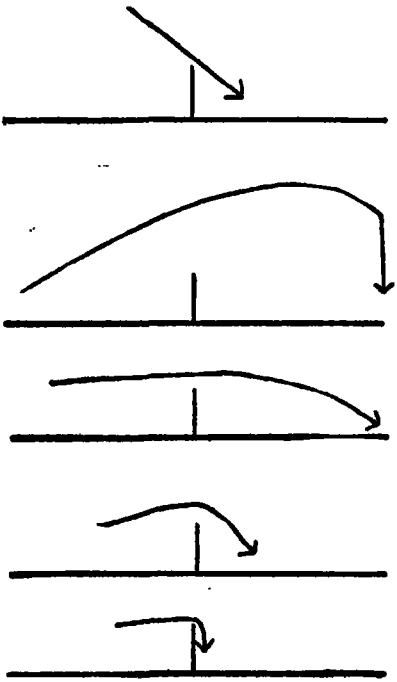


Scoring: The ladies' singles game consists of .... points, ladies' doubles .... points, and mixed doubles .... points. In doubles, when the score is "13 all," the side which first reached 13 has the option of "setting" the game to ...., and when the score is "14 all," the game may be set to ..... In the singles game, when the score is "9 all," the game may be set to ...., and when the score is "10 all," the game may be set to ..... A match consists of .... games.

Differences Between Doubles and Singles: The side alleys are used only in the .... game. In the .... game, the long service line is the same as the back boundary line. In the .... game, the service area is wide and short, compared to the .... game, where the service area is narrow and long.

Take a pencil and shade the area in the diagram that represents the doubles service court. Then do likewise for the singles service court.

Strokes (Diagrams indicate the line of flight of the bird)



Name:
Uses:

How to return it:

Name:
Uses:

How to return it:

Name:
Uses:
How to return it:

Name:
Uses:

How to return it:

Name:
Uses:
How to return it:

Equipment—How should racquets and shuttlecocks be cared for?

Terminology—Define the following terms:

Fault:
Home plate:
Side out:
Slung:
Hand out:
Inning:

Systems of Court Coverage in Doubles:

Name ..... Advantages .....
Disadvantages .....

Name ..... Advantages .....
Disadvantages .....

Name ..... Advantages .....
Disadvantages .....

Self-Inventory:

1. How many of the following strokes do I actually use in a game?
   a. short serve
   b. long serve
   c. clear
   d. hairpin drop shot
   e. cross court drop shot
   f. long drop shot
   g. smash
   h. drive

2. What strategy do I employ?
   a. Placement of serve to opponent's backhand.
   b. Placement of all strokes to open areas, thus forcing opponent to move.
   c. Use of drop shot to lead opponent into setting bird up for a smash.
   d. Placing strokes to the vulnerable backhand rear corner.
   e. Feint.
   f. Change of pace.

Part I. Rules Chart. Fill in the blank spaces in the chart below and on another piece of paper, make a similar form, extending the chart. Use this chart in reviewing for examination. The first one has been filled in as an example.

Description of what happened	Foul, violation or legal play	If foul, what type	Name	Penalty
0. Forward in act of shooting pushes guard away from her by placing the ball against guard's chest and pushing with it.	Foul	Personal	Charging	1 free throw
1. One player in possession of the ball is guarded between two players and is unable to pass the ball before the three second time limit has been reached.				
2. Player, by use of personal contact, impedes the progress of an opponent who has started to advance the ball by means of a bounce or a juggle.				
3. Player restricts the fredeom of movement of an opponent without the ball by disregarding the ball and shifting her position as the opponent moves.				
4. Player in act of shooting is bumped from the rear by an opponent. Basket is missed.				
5. As above (4) except basket is made.				
6. Players are jumping for rebound. Red player gets both hands on the ball; Blue player gets one hand on the ball slightly later, and thinking she has legally tied the ball keeps hand on it.				
7. Player snatches the ball out of the hands of an opponent who has it legally in her possession.				
8. Player fumbles the ball in catching it, recovers it from the floor, and then bounces it.				
9. On toss up, one of the players taps the ball twice in succession.				
10. Team has had three time outs. Captain requests a fourth time out.				

Figure 41. Sample Work Sheet for Basketball

was dropped. This particular question could probably be made more difficult by removing the clause in the stem, and adding those three strokes at various places in the responses.

A question which proved difficult for the same group (difficulty rating was 93) was as follows:

> Girls' volley ball. The ball is served by team A to the center back player on team B. What should the center back do?
>
> (1) Set the ball up to herself and spike it.
> (2) Set the ball up to a center row player.
> (3) Set the ball up to herself and pass to a spiker in front row.
> (4) Set the ball up to herself and pass to a short player in the front row.
> (5) Return the ball to the rear of team A's court.

The greatest difficulty encountered was in selecting between the second and fourth items, although the errors in selection were distributed well enough to cause all parts to function. There are several ways in which the question could probably be made easier: eliminate item 2, or indicate the need for taking the spin off the ball by setting it up to self (do this in the stem), or add the word "immediately" or "directly" to item 5.

Care must be taken in attempting to revise a question to keep the intended purpose of the question in mind and not to make it easier by the simple device of throwing in responses that no one will select. A question with a long and involved stem can be clarified sometimes by breaking it into two sentences.

TESTS AS TEACHING DEVICES

Sometimes teachers will want to use tests for instructional purposes, with no desire to use them as a partial basis for grading. One such device is the badminton worksheet presented here. This and similar worksheets can be used while students are resting between periods of play, in connection with the showing of movies, or can be filled in outside of class time. A master sheet with correct answers can be posted, if the teacher does not have the class time to go over the answers with the students. The work-

sheets may be retained by the students, and they may make additions to them, to use in reviewing for the course examination. The experienced teacher will have no difficulty in preparing similar worksheets for his classes, selecting those things which he considers most important to stress.

The badminton worksheet illustrates a justifiable use of the completion type of exercise. The purpose here is to provide the student with the correct information, not to test him. An effort should be made to encourage the student to make comparisons and to think for himself, rather than merely transferring answers directly from a text or lecture. If the worksheet is to be corrected by the teacher the form should be changed with blank spaces placed in the margin.

Portions of worksheets * in other activities follow, illustrating one method of handling the content.

FOLK (OR COUNTRY) DANCE

Characteristics of specific dances:

	Name of dance	*No. in dance*	*Origin*	*Steps*	*Floor pattern*
1					
2					
3					
.					
.					
.					
.					
10					

TENNIS

A. Singles

1. Draw a diagram of the court and indicate the "vital area" and "no man's land."
2. Explain the reason for returning to a position behind the center of the baseline after each stroke.

* These illustrations are adapted from worksheets prepared for use in college classes, University of Minnesota, by staff members: Folk or Country Dance, Mary Virginia Gardner and Elizabeth Kratz; Tennis, Virginia Pettigrew; Archery, Eloise Jaeger and Catherine Snell; Badminton and Basketball, Esther French; Posture and Conditioning, Ellen Kelly.

B. Doubles

1. Explain the parallel system of court coverage.
2. List the disadvantages of the "up and back" system of court coverage.

ARCHERY

A. List the common errors made in shooting, the result, and the correction.

	Error	*Result*	*Correction*
1.			
2.			
3.			
4.			
5.			
6.			

B. Describe the following types of competition (on a separate sheet of paper):

1. Columbia
2. American
3. National
4. York
5. Clout shoot
6. Archery golf
7. Roving
8. Flight shoot

C. Supply answers on a separate sheet:

1. What are the causes of the arrow falling off the hand?
2. How should arrows be removed from the target?
3. How should the bow be bent?
4. How should the arrow be nocked?
5. What are some of the causes of injury?
6. Draw a diagram of an arrow, naming the parts.

POSTURE AND CONDITIONING

A. Describe or name exercises for various parts of the body and various purposes.

1. Improve circulation (warm-up and flexibility).
2. Abdominal strength.
3. Arm and shoulder strength.
4. Upper back strength.
5. Foot strength and correct use.

6. Leg strength.
7. Relaxation.
8. Body alignment.

(This same type of knowledge can be obtained by listing the exercises vertically and the purposes horizontally, thus providing a checkerboard diagram for check marks.)

B. Supply answers on a separate sheet.

1. Standing
   a. How should one's body weight be distributed over the feet?
   b. When viewed from the side, which landmarks should be in a vertical, straight line over the middle of the foot?
   c. When viewed from the back, which landmarks should fall in a straight, vertical line centered between the feet?
   d. What should you do in testing your posture against a wall?
2. Sitting
   a. List the characteristics of a properly fitting chair and a desirable posture for active work and study.
   b. How can fatigue and backstrain be minimized in sedentary jobs?
   c. Describe an attractive and efficient form in getting into and out of a chair.
   d. What is a good sitting position for recreational reading?
3. Walking and running
   a. How can the arches be protected from undue strain in walking and running?
   b. What are the factors which add or detract from the appearance in walking?
4. Lifting
   a. What form should be used in lifting heavy loads with a minimum of fatigue and without danger of back injury?
   b. How can greatest force be applied in lifting?

When tests are to be used for teaching purposes only, it is possible to make use of many of the forms that can not be used in objective tests because of the difficulties involved in scoring. Some of the forms that have proven valuable are definitions, short essays, and "listings." Be sure to keep the purpose in mind and avoid the use of teaching types of questions except in those examinations which are aimed primarily as teaching devices.

BIBLIOGRAPHY

1. Bird, Charles, and Andrew, Dorothy M.: "Concerning the Length of the New Type Examination," *Journal of Educational Psychology*, 27, 1936, p. 641
2. ——: "Comparative Validity of New Type Questions," *Journal of Educational Psychology*, 28, 1937, p. 241
3. French, Esther: "The Construction of Knowledge Tests in Selected Professional Courses in Physical Education," *Research Quarterly*, 14, Dec. 1943, p. 406
4. Hawkes, Herbert E., Lindquist, E. F., and Mann, C. R.: *Construction and Use of Achievement Examinations*. Houghton Mifflin Company, Boston, 1936
5. Hewitt, Jack E.: "Comprehensive Tennis Knowledge Test," *Research Quarterly*, 8, October 1937, p. 74
6. Lindquist, E. F., and Cook, Walter W.: "Experimental Procedures in Test Evaluation," *Journal of Experimental Education*, 1 1933, p. 163
7. Pullias, E. V.: "Problems in Evaluation of Responses," *Education Methods*, 18, 1938, p. 75
8. Richardson, M. W., Russell, J. T., Stalnaker, J. M., and Thurstone, L. L.: *Manual of Examination Methods*. University of Chicago Press, 1933
9. Ruch, G. M.: *Improvement of the Written Examination*. Scott Foresman and Company, Chicago, 1924
10. Rugen, Mabel E., and Saurborn, Jeanette: *Physical Education Teaching Manual*. Edwards Bros. Inc., Ann Arbor, 1936, Chapters 10 and 11
11. Scott, M. Gladys: "Achievement Examinations for Elementary and Intermediate Swimming Classes," *Research Quarterly*, 11, May 1940, p. 100
12. ——: "Achievement Examinations for Elementary and Intermediate Tennis Classes," *Research Quarterly*, 12, March 1941, p. 40
13. ——: "Achievement Examinations in Badminton," *Research Quarterly*, 12, May 1941, p. 242
14. Snell, Catherine: "Physical Education Knowledge Tests," *Research Quarterly*, 6, October 1935, Fundamentals, p. 79; Hockey, p. 86.
15. ——: "Physical Education Knowledge Tests," *Research Quarterly*, 7, March 1936, Basketball, p. 79; Soccer, p. 76; Volley Ball, p. 73
16. ——: "Physical Education Knowledge Tests," *Research Quarterly*, 7, May 1936, Baseball, p. 87; Golf, p. 77; Tennis, p. 84
17. Swineford, Frances: "Validity of Test Items," *Journal of Education Psychology*, 27, 1936, p. 68
18. Trumbull, Katharine S.: "Planning for Results," *Journal of Health and Physical Education*, 15, January 1944, p. 19

9. Simple Statistical Proceedings

In the belief that it is important for all teachers of physical education to have at least a speaking acquaintance with simple statistical procedures, this chapter is included. Many excellent textbooks are available for teaching methods of statistical computation.[2, 4, 5] Texts which point out the limitations of the various devices are recommended, because it seems to the authors that there is danger in the misuse of techniques by persons with only a limited understanding of the underlying assumptions that are made in the development of each technique. The importance of applying logic to research cannot be over emphasized.

THE FREQUENCY DISTRIBUTION

Let us suppose that you have given a test to a group of individuals and have one score for each individual. To get any generalized concept of the performance of the group and of the meaning of each score, it is necessary to arrange them in some manner. To apply the example to physical education, take the scores obtained from forty-three college girls on a badminton clear test (see p. 51 for the test description). The test was given at the end of seven periods of instruction. A perfect score on the test is 100.

TABLE XV

SCORES OF FORTY-THREE STUDENTS ON THE BADMINTON CLEAR TEST

45	6	75	60	78	47	65	65	63
32	22	35	56	74	54	72	64	72
28	39	47	54	73	59	78	64	63
69	86	49	85	69	64	79	60	
60	80	52	80	84	61	66	62	

[4, 5] See references in bibliography at end of this chapter.

The raw scores, as presented in Table XV, are not very meaningful or helpful in evaluating the relative performance of any individual within the group. It is apparent that no student made a perfect score, and also that the highest score made was 86 and the lowest score 6; this arrangement does make it possible to obtain the range. (The range is the difference between the highest and the lowest score in the series.)

The scores can be interpreted better if all the possible scores within the range are listed in order of size, and then the number of times each score was made is recorded, as has been done in Table XVI.

TABLE XVI

SIMPLE FREQUENCY DISTRIBUTION OF SCORES ON BADMINTON CLEAR TEST

(Intervals of one unit)

S	F	S	F	S	F	S	F	S	F
86	1	69	2	52	1	35	1	18	
85	1	68		51		34		17	
84	1	67		50		33		16	
83		66	1	49	1	32	1	15	
82		65	2	48		31		14	
81		64	3	47	2	30		13	
80	2	63	2	46		29		12	
79	1	62	1	45	1	28	1	11	
78	2	61	1	44		27		10	
77		60	3	43		26		9	
76		59	1	42		25		8	
75	1	58		41		24		7	
74	1	57		40		23		6	1
73	1	56	1	39	1	22	1		
72	2	55		38		21			N = 43
71		54	2	37		20			
70		53		36		19			

The more frequently occurring scores now stand out, the points of concentration of scores are apparent, and the number of scores between any two given points can be quickly secured by simple addition. (The table is usually arranged in a linear form, instead of in the separate five columns shown here, an arrangement to conserve space. When in linear form the graphic distribution of scores is more apparent.) But Table XVI needs to be condensed into "classes" of scores, for convenience in handling. Each "class" or

"step-interval" will include all the records of scores within the limits of that interval. The size of the step interval will depend, somewhat, upon the use to be made of the data. There is seldom any need for fewer than twelve intervals or more than twenty. Dividing the range by fifteen will give a quick approximation of the size of the interval. Example, range: 86 (highest score) minus 6 (lowest score) or 80; 80 divided by 15 is 5.3, or in round numbers, 5. For ease and accuracy in tabulation, an interval of 1, 2, 3, 5, 7, 10, 15, or any higher multiple of 5 is preferred. The increase or decrease in the number of classes by one or two is not important. Table XVII shows the grouped frequency distribution of the same data as used in Tables XV and XVI, using intervals of five units.

TABLE XVII

GROUPED FREQUENCY DISTRIBUTION OF SCORES ON BADMINTON CLEAR TEST

(Intervals of Five Units)

S	Tab	F
83-87	///	3
78-82	~~////~~	5
73-77	///	3
68-72	////	4
63-67	~~////~~ ///	8
58-62	~~////~~ /	6
53-57	///	3
48-52	//	2
43-47	///	3
38-42	/	1
33-37	/	1
28-32	//	2
23-27		0
18-22	/	1
13-17		0
8-12		0
3- 7	/	1
		N = 43

This table could be made more compact by increasing the size of the step-interval. By using an interval of 25 units, for example, only four classes would be needed. But such an increase in the size of the unit would mean a larger loss in the identity of the original scores. If an interval of 25 were used, over half of the scores would

fall in one interval (63-87), thus hiding most of the characteristics of the original distribution. If only a very rough picture of the distribution of scores is needed, then a very broad interval may prove satisfactory. If any detailed study is to be made, or if high precision in description is desirable, then the interval used should be small.

The steps in constructing a grouped frequency distribution, then, are as follows:

1. Prepare a data sheet with the three headings, as shown in Table XVII. (The abbreviation S is for scores; Tab for tabulation of tally marks; and F for frequencies, or the number of times that a score occurs within that interval.)

2. Determine the range and divide by 15. (Carry the results to only one decimal place.)

3. Select from the following preferred list the number nearest the quotient secured in Step 2: 1, 2, 3, 5, 7, 10, 15, or any higher multiple of 5.

4. Write the limits of each interval, in descending order, in the first (S) column of the table, beginning at the top with the interval containing the highest score. Determine these limits as follows:

   (a) When the number of units in the interval is an even number, the lower limit of each interval should be a multiple of this number.

   (b) When the number of units in the interval is an odd number, find the multiple of this number that is nearest to the highest score in the series. (In the example, using an odd number for the interval (5), with the highest score being 86, the nearest multiple is 85.) Select the limits of the interval so that this multiple is the middle score in the interval. Thus the midpoints of all the intervals will be a multiple of the interval unit size. (See example, Table XVII.) The reason is for convenience only. Because of the loss of identity of the original scores in any grouped frequency distribution, it is necessary in later computations to use the midpoint to represent the value of the scores in the interval. It is convenient, then, to have the midpoint be a whole number, rather than a decimal value, and also

to have it be a multiple of the step-interval. Note that the use of an interval containing an odd number of units results in a more convenient midpoint, as the midpoint of an even interval will be a decimal value.

5. Tabulate the scores, by placing a tally mark opposite the appropriate interval.

6. Record the number of tally marks opposite each interval in the frequency column and add, as an immediate check on the accuracy of tabulation. (See N, the symbol for the total number of cases, in Table XVII.)

The frequency distribution has been described in detail because it is the basic step for practically any statistical procedure.

CONVERTING RAW SCORES INTO LETTER GRADES

Raw scores can be quickly transferred into letter grades, when placed in a frequency distribution. Let us suppose that you have decided to base the grades in a particular course as follows: one-third to be determined by skill test scores, one-third by knowledge test scores, and one-third by subjective estimate. If your school uses letter grades, a percentage system such as 7% A, 24% B, 38% C, 24% D, and 7% E (or failure) has probably been established. The limits of each letter grade for the various scores in Table XVI can be determined by multiplying N (43) by each of the respective percentages, rounding the figures. Seven times 43 is 3.01 or 3, so the value of A is assigned to the top three scores. If a plus and minus system is used, the score 86 would have the value of A +, and the score 84 the value of A —. Twenty-four times 43 is 10.32 or 10. The scores falling between 70 and 80 are given B grades. Thirty-eight times 43 is 16.34 or 16. Counting sixteen down, the C's will include scores from 69 to 57, inclusive. The number of D's will be the same as the number of B's, since the percentage is the same, so the next ten scores are assigned the value of D. This includes the scores through 33, and four scores remain. This is not an error but is the result of the rounding of numbers. The question now is shall we assign a value of D to the score 32 or the value of E. Since 32 is closer to 35 than to 28, it seems fair

to place it in the D group, and extend the interval to include 32, as shown in the table below.

			F
81 and up	:	A	3
70-80	:	B	10
57-69	:	C	16
32-56	:	D	11
31 and below	:	E	3
			N = 43

When this is done from a grouped frequency distribution, a few more such complications may occur. For example, try using Table XVII, which has the same data, but grouped in intervals of five units. The top (A) interval now becomes 83 and up. The B's will include the five recorded in the 78-82 interval, the three in the 73-77 interval, and it will be necessary to include the four in the 68-72 interval. This gives you 12 B's. It is obvious that if you start the C's by counting sixteen cases, you will again find yourself giving too many and will end without any failures. To prevent this, subtract the extra number of B's (2) from the total number of C's (16) and give just 14 C's. Fortunately, this break comes at the end of an interval, so scores 58-67 inclusive are assigned C's. You now start with a clean slate, to assign 24% of the cases, or 10, to the D category. The D's then include scores of 33-57. This time you have just three remaining scores, for the failures. Your scoring table now is as follows:

			F
83 and up	:	A	3
68-82	:	B	12
58-67	:	C	14
33-57	:	D	10
32 and below	:	E	3

Sometimes it is simpler and gives a more equitable distribution of grades if the procedure is changed. After computing the correct number of cases for 7% and counting them off at the top, take the other end of the distribution and count up in reverse fashion to determine the limits of E. Likewise, when the limits for the 24% B's have been set, count up from the E's to determine the limits for the corresponding set of D's. This may leave slightly more or

less than 38% in the center for C's, but this represents the average group and it is probably better to have it overloaded than to have one of the other categories altered.

It is also necessary to make intelligent modifications in the application of the percentage system. If the upper scores fall far short of expectations or fail to measure up to known standards for such a group, the teacher may reduce the number of A's or eliminate them entirely. In this instance the extra number of cases will be added on to the B or C category, depending upon where there is the greatest concentration of scores. Also, on the low end of the distribution the poorest scores may not be really poor in terms of the teacher's expectations for the class, in terms of known standards of performance, or in terms of actual difference from average or best scores for that particular group. Under these circumstances, the teacher would probably not give any E's, but increase the number of D's or C's.

Also, if there are long gaps in the distribution they may form more natural and just divisions between categories than those defined by the percentage system. For example, by comparing Tables XVI and XVII, the latter shows a little more clearly that 28 is closer to 32 than to 22. Working from Table XVII the teacher would be much more apt to assign only 2 E's, to scores 22 and 6. This apparently large gap between 28 to 22 is partly the result of the limits defined for the step-interval; but the actual assignment of 28 to a D or an E must finally be determined by a subjective estimate of how poor 28 really is. By looking at the total distribution and the difference between it and the maximum score it would appear really poor, but so would scores 32, 35 and 39. It is possible that in some cases the teacher might assign E's to 39 and all scores below it, though this would probably be too drastic if it is a final term grade.

The larger the number of cases in the distribution the fewer and shorter will be the gaps between scores. If the frequency distribution shown in Table XVII is kept and scores added to it for a number of classes or seasons, and the limits are adjusted each time, more stable divisions will gradually evolve.

AVERAGING LETTER GRADES

It is often necessary to average several letter grades to obtain one grade for each student. If you are basing one-third of the grade on skill test scores, one-third on knowledge test scores, and the other one-third on a rating of the student in the activity itself, you may have as many as three or more grades to average, depending on the amount of testing that you have done. These may be further complicated by the plus and minus system. Points can be assigned to each grade, totalled, and then divided by the number of grades, in a system somewhat similar to the honor point system. Suppose Jane Doe has the following record: A—, C+, D, B—, E, C, B+, C—. By use of a point value table, these scores can be converted into a total of 43 points. When divided by 8 this gives a point score of 5. This is equivalent to a grade of *C*. The point value table appears below.

TABLE XVIII

POINT VALUE TABLE FOR AVERAGING LETTER GRADES

A+ : 12	B+ : 9	C+ : 6	D+ : 3	E : 0
A : 11	B : 8	C : 5	D : 2	
A— : 10	B— : 7	C— : 4	D— : 1	

This table may also be used on the knowledge and rating phases of her final grade. Let us suppose that the eight scores above were skill test scores, so Jane now has a single score of C for skill, as measured by the eight tests. If she made a D on the final knowledge test, the only one given, and was rated a B player, her score can be determined without use of the above table. But if her knowledge score was A+ and her rating D, with skill score C, the table may be needed to obtain a total of 19 points, divided by 3, with the result of 6.3, or an average grade of C+.

MEASURES OF CENTRAL TENDENCY

Before proceeding with the discussion of methods of converting raw scores into comparable scores (scores that can be averaged to obtain a single grade) measures of central tendency

need to be considered. This discussion will be limited to the two most commonly used in handling physical education test data, namely, the mean and the median.

The mean (M) does not involve a new concept; it is simply the arithmetical average. It is obtained by adding all of the scores and dividing by the number of individuals. It is based, then, on every score in the distribution. The mean being a mathematically derived value, may be used in further computations.

The median (Mdn) is the middle measure in a series in which all of the measures have been arranged in the order of their size. It may be quickly computed from a frequency distribution by dividing the N by two and then counting up that number of frequencies. When the number of cases is small and one or more cases deviate markedly, the median will give a better representation of the typical than will the mean, which is affected by all the scores.

The mean and the median will be computed for the data presented in Table XVII, Finding the Arithmetic Mean of a Grouped Frequency Distribution.

The mean can be found by the laborious method of summing all the original scores, as presented in Table XV, and then dividing by N. When the number of cases is small, as in this example, that can be done without great effort. But frequently it must be computed from a grouped frequency distribution containing many cases. The steps involved are as follows:

1. Select as an arbitrary reference point the interval which you think is most likely to contain the actual mean.

2. Express each interval above and below the interval containing the arbitrary reference point (A.R.) as a deviation. Record in the d column, as in the illustration. All the deviations below the A.R. must be preceded by a negative sign.

3. Multiply the frequency in each interval by the corresponding d value, and record the products in the column headed fd. (Note that all those below the A.R. will have a negative sign.)

4. Add the positive products, then the negative products, and add these sums *algebraically*.

5. Divide this result obtained in Step 4 by N. This represents the correction to the A.R., expressed in interval units. To make

the correction complete in score units, this quotient must be multiplied by the size of the interval, (5, in the illustration).

6. Add this product algebraically to the A.R., which for this purpose is the midpoint of that interval. (Add if positive, subtract if negative.)

When studying Table XIX for the computation of the mean, disregard the column headed cf.

TABLE XIX

COMPUTATION OF MEAN AND MEDIAN FOR FREQUENCY DISTRIBUTION ON BADMINTON CLEAR TEST

	f	*d*	*fd*	*cf*	
83-87	3	5	15	43	$\Sigma fd = 60 - 56 = 4$
78-82	5	4	20	40	
73-77	3	3	9	35	$\frac{\Sigma fd}{N} = \frac{4}{43} = .093$ or .09 (rounded)
68-72	4	2	8	32	
63-67	8	1	8	28	
			+60		
A.R. 58-62	6	0		20	.09 is the correction in interval units
53-57	3	−1	−3	14	
48-52	2	−2	−4	11	
43-47	3	−3	−9	9	
38-42	1	−4	−4	6	
33-37	1	−5	−5	5	Correction in score units = $5 \times .09 = .45$
28-32	2	−6	−12	4	
23-27	0	−7		2	
18-22	1	−8	−8	2	
13-17	0	−9		1	Mean = A.R. + correction in score units = 60 + .45 = 60.45
8-12	0	−10		1	
3-7	1	−11	−11	1	
			−56		

FINDING THE MEDIAN OF A GROUPED FREQUENCY DISTRIBUTION

The median in a simple frequency distribution, as shown in Table XVI, can be quickly ascertained by dividing N by two and finding that point on the scale above or below which half of the frequencies lie. In the example in Table XVI the median would fall at 63. To compute it from a grouped frequency distribution, the fact that we have lost the identity of the original scores

necessitates a somewhat more complicated procedure. The first step is to add a cumulative frequency column to the frequency distribution. See the column headed *cf* in the illustration in Table XIX. This is done by totalling the frequencies, interval by interval from the bottom up. The median is by definition the middle score in the distribution, and is located by the formula $\frac{N+1}{2}$ if N is an odd number. (If N is an even number, then the median is computed as half way between the two central scores; i.e., if N is 44, the median would be half way between the twenty-second and twenty-third scores.) In this case N is 43, therefore the median is $\frac{43+1}{2}=22$. It is noted immediately that the twenty-second score falls in the interval 63-67. The next step is to subtract from 22 the number in the cumulative frequency column just below (20). Then, divide this difference (2) by the frequency in this interval (8). Multiply this quotient ($\frac{2}{8}=.25$) by the size of the interval (5). Add this product (1.25) to the lower limit of the interval (63 + 1.25 is 64.25). The median then is 64.25, or in round numbers 64.

The effect of the one extremely low score on the mean is now apparent when a comparison is made between the mean and the median. The mean is 60.45 and the median 64.25. In a normal distribution, the mean and the median would coincide. The one low score does not affect the median any more in its actual location in the distribution than it would have if it had varied several intervals in either direction from its actual location. The median, in this case, gives a better description of the typical individual's score. The mean, because of its use in other computations, is important.

MEASURES OF DISTRIBUTION

When the mean or median has been determined as a measure of central tendency it tells only a single value which is to be used to represent the total distribution. This measure does not tell anything about how much the scores deviate from it or with what frequency they approach it. For example, if a group of sixth grade

boys has taken the standing broad jump and you know that the average jump was 66 inches you might assume that many scored near that point. However, you do not really know about the performance of the total group until you have some measure of the distribution or spread. Or in the previous example of the badminton clear test, if you know only that the mean is 60.45 you are entirely unaware of the great variability of scores presented in that distribution.

The distribution around the median is usually interpreted by use of the quartile (Q), and that with the mean by use of the standard deviation (S.D. or σ). The quartile simply splits each half designated by the median. The first and third quartile points are located by exactly the same process as the median was.

The standard deviation is computed from a grouped frequency distribution as in Table XIX. Add to that table another column marked fd^2. The values in this column are obtained by multiplying those in the fd column by those in d. The computation then proceeds as in Table XX.

TABLE XX

COMPUTATION OF STANDARD DEVIATION FOR FREQUENCY DISTRIBUTION ON BADMINTON CLEAR TEST

	f	*d*	*fd*	fd^2	
83-87	3	5	15	75	$\Sigma fd^2 = 542$
78-82	5	4	20	80	
73-77	3	3	9	27	$\frac{\Sigma fd^2}{N} = \frac{542}{43} = 12.6$
68-72	4	2	8	16	
63-67	8	1	8	8	$c = .09$ (computed in Table XIX)
58-62	6	0			
53-57	3	— 1	— 3	3	$c^2 = .008$
48-52	2	— 2	— 4	8	
43-47	3	— 3	— 9	27	S.D. $= \sqrt{\frac{\Sigma fd^2}{N} - c^2} \times$ size of step-interval
38-42	1	— 4	— 4	16	
33-37	1	— 5	— 5	25	
28-32	2	— 6	— 12	72	$= \sqrt{12.6 - .008} =$
23-27	0	— 7			
18-22	1	— 8	— 8	64	$\sqrt{12.592} = 3.6$
13-17	0	— 9			
8-12	0	— 10			3.6 × size of step-interval =
3- 7	1	— 11	— 11	121	3.6 × 5 = 18.
				542	

The fd^2 column is added ($\Sigma fd^2 = 542$) then divided by N $\left(\frac{\Sigma fd^2}{N} = 12.6\right)$. The correction, computed for the arbitrary reference point in obtaining the mean, is squared and then subtracted from the last quotient ($12.6 - .008 = 12.592$). The square root of this number is then determined to complete the last step in the formula ($\text{S.D.} = \sqrt{\frac{\Sigma fd^2}{N} - c^2} = 3.6$). This gives the standard deviation in terms of step-intervals. To interpret in terms of raw scores multiply by the size of the step-interval ($3.6 \times 5 = 18.0$).

See Figure 4, p. 27 for the illustration of percentages within each S.D. distance in the distribution. The S.D. is small if the total range is small and the cases cluster closely around the mean. The S.D. becomes larger as the range increases and the cases spread. The proportion of cases between any two S.D. points in the distribution remain the same as long as the curve maintains characteristics resembling the normal curve. (See Figure 4) If the distribution becomes too dissimilar to the normal curve the S.D. should not be used.

The percentile rankings of scores may be computed in connection with either the median or the mean. However, the percentiles are usually used for interpretation of raw scores, particularly for comparing abilities measured by different tests. The scores range from zero to the 100th percentile. A score having a percentile ranking of 34 is better than 34 percent of the cases and poorer than 66 percent of them. (See any statistics text for the technique of computation.)

CONVERSION OF RAW SCORES INTO COMPARABLE VALUES

The T-scale can be very easily constructed on raw scores collected on any class or group of similar classes. There should be a minimum of fifty students represented in the scale, preferably a hundred or more. For this reason, if there are two or three sections of the same age group working on the same activity, it is better to combine the scores from all classes into a single distribution, rather than using a separate one for each sec-

tion. This combination puts the standards represented by the scale on a school basis rather than entirely on their own particular group.

The T-scales presented in previous chapters are based on several sections within the same school or on similar classes for several schools. The number of subjects is always noted and usually the source of the subjects.

The T-scale is based on the characteristic distribution of cases in a *normal curve* or in one which approximates that curve. The teacher or person who constructs the scale should understand something about that curve, viz., the two ends are symmetrical, the scores concentrate heavily at the center, the two extremes are at approximately equal distance from the center and the frequencies near the extremes fall off very rapidly. See Figure 4 for the frequencies within each standard deviation. The T-scale is based directly on the standard deviation. The middle score is arbitrarily assigned a T-score of 50 and since the mean is at the center of the curve the two are identical. The normal curve extends three standard deviations in each direction. By assigning 10 T-scores to each standard deviation the scale then must range from approximately 20 to 80. Since the scores are massed near the center there will be many who receive a score in the 40's or 50's. There will be considerably fewer who receive 30's or 60's, and comparatively few who receive 20's and 70's.

Such a scale is easily explained to students even of high school age. They are familiar with the concept of the average. They will understand the following explanation. Scores near fifty are about average, those considerably above or below 50 represent good or poor performance respectively. In addition to enabling the student to understand his relative ability on the single test he may also make a similar comparison of his performance on several tests.

From the standpoint of the teacher or test administrator it is possible to add or average T-scores for composite ratings. This is a short cut to a score on a test battery, and is sometimes the only means of obtaining a single score on a series of tests.

TABLE XXI

COMPUTATION OF A T-SCALE FOR 200 CASES ON A MOTOR ABILITY BATTERY

Step-interval	*Tallies*	*f*	*t*	½	*t*+½	%	*T-Score*
156-up	/	1	0	.5	.5	.250	78
154-155	/	1	1	.5	1.5	.750	74
152-153	/	1	2	.5	2.5	1.250	72
150-151	//	2	3	1.0	4.0	2.000	71
148-149	///	3	5	1.5	6.5	3.250	68
146-147	////	4	8	2.0	10.0	5.000	66
144-145	///	3	12	1.5	13.5	6.750	65
142-143	///	3	15	1.5	16.5	8.250	64
140-141	~~++++~~	5	18	2.5	20.5	10.250	63
138-139	~~++++~~	5	23	2.5	25.5	12.750	61
136-137	~~++++~~ //	7	28	3.5	31.5	15.750	60
134-135	~~++++~~ /	6	35	3.0	38.0	19.000	59
132-133	~~++++~~ ///	8	41	4.0	45.0	22.500	58
130-131	~~++++ ++++~~	10	49	5.0	54.0	27.000	56
128-129	~~++++ ++++~~ ////	14	59	7.0	66.0	33.000	54
126-127	~~++++ ++++~~ ///	13	73	6.5	79.5	39.700	53
124-125	~~++++ ++++ ++++~~ /	16	86	8.0	94.0	47.000	51
122-123	~~++++ ++++~~ //	12	102	6.0	108.0	54.000	49
120-121	~~++++ ++++~~	10	114	5.0	119.0	59.500	48
118-119	~~++++~~ ////	9	124	4.5	128.5	64.250	46
116-117	~~++++ ++++~~	10	133	5.0	138.0	69.000	45
114-115	~~++++ ++++~~ /	11	143	5.5	148.5	74.250	43
112-113	~~++++~~ ////	9	154	4.5	158.5	79.250	42
110-111	~~++++~~ ///	8	163	4.0	167.0	83.500	40
108-109	~~++++~~ //	7	171	3.5	174.5	87.250	39
106-107	~~++++~~ //	7	178	3.5	181.5	90.750	37
104-105	~~++++~~	5	185	2.5	187.5	93.750	35
102-103	///	3	190	1.5	191.5	95.750	33
100-101	//	2	193	1.0	194.0	97.000	31
98- 99	//	2	195	1.0	196.0	98.000	29
96- 97	/	1	197	.5	197.5	98.750	28
94- 95	/	1	198	.5	198.5	99.250	26
93-down	/	1	199	.5	199.5	99.750	22

N = 200

The steps in construction of a T-scale follow:

1. Make a frequency distribution of scores to be used. (This is the same procedure as outlined above except that in this case the size of the step-interval must be small. If intervals are too large there will be too much massing of raw scores and the T-scale will be broken rather than continuous. In other words, the scale will not discriminate between performance of different persons with similar scores.)

2. Total the frequencies in the column marked *f*.

3. Count the total of all frequencies above each interval and put the total in column *t*. For example, there is nothing above the top interval, therefore, the *t* column reads 0 in the first interval; in the successive intervals of *t*, the value will be the sum of *f* and *t* of the interval above. As a check on accuracy, the sum of *f* and *t* of the last interval always should be equal to N.

4. Divide *f* values for each interval by 2. Therefore, this column is called the ½ column.

5. Add the *t* and ½ columns for each interval; label the column accordingly $t + ½$. (The purpose of this step is to find the number of frequencies above the midpoint of each interval. This is done because the midpoint of the interval is always considered as representative of all cases in the interval. This is another reason for keeping the size of the step-interval small.)

6. Divide the $t + ½$ column by N and multiply by 100. This column is called the % column. (Carry the percentages to at least three decimal points.)

7. Read in Table XXII the standard deviation value corresponding to each percentage value. This is the T-score and is inserted in a column so labelled. (Use the T-score to the nearest whole number rather than in decimals.)

8. Eliminate the computation leaving only the values for the step-interval and the corresponding T-score. If this is entirely for your own use the quickest way is to simply fold the sheet over so that the T-score column is beside the step-interval column. If you wish a more permanent form or wish to post it, copy only those two columns on another sheet, perhaps including T-scales for several tests on the same sheet for greater convenience. (See p. 40.)

TABLE XXII

Conversion of Percentages into T-Scores

T-Scores *	*Per cent*	*T-Scores*	*Per cent*	*T-Scores*	*Per cent*	*T-Scores*	*Per cent*
0	99.999971	25	99.38	50	50.00	75	0.62
0.5	99.999963	25.5	99.29	50.5	48.01	75.5	0.54
1	99.999952	26	99.18	51	46.02	76	0.47
1.5	99.999938	26.5	99.06	51.5	44.04	76.5	0.40
2	99.99992	27	98.93	52	42.07	77	0.35
2.5	99.99990	27.5	98.78	52.5	40.13	77.5	0.30
3	99.99987	28	98.61	53	38.21	78	0.26
3.5	99.999883	28.5	98.42	53.5	36.32	78.5	0.22
4	99.99979	29	98.21	54	34.46	79	0.19
4.5	99.99973	29.5	97.98	54.5	32.64	79.5	0.16
5	99.99966	30	97.72	55	30.85	80	0.13
5.5	99.99957	30.5	97.44	55.5	29.12	80.5	0.11
6	99.99946	31	97.13	56	27.43	81	0.097
6.5	99.99932	31.5	96.78	56.5	25.78	81.5	0.082
7	99.99915	32	96.41	57	24.20	82	0.069
7.5	99.9989	32.5	95.99	57.5	22.66	82.5	0.058
8	99.9987	33	95.54	58	21.19	83	0.048
8.5	99.9983	33.5	95.05	58.5	19.77	83.5	0.040
9	99.9979	34	94.52	59	18.41	84	0.034
9.5	99.9974	34.5	93.94	59.5	17.11	84.5	0.028
10	99.9968	35	93.32	60	15.87	85	0.023
10.5	99.9961	35.5	92.65	60.5	14.69	85.5	0.019
11	99.9952	36	91.92	61	13.57	86	0.016
11.5	99.9941	36.5	91.15	61.5	12.51	86.5	0.013
12	99.9928	37	90.32	62	11.51	87	0.011
12.5	99.9912	37.5	89.44	62.5	10.56	87.5	0.009
13	99.989	38	88.49	63	9.68	88	0.007
13.5	99.987	38.5	87.49	63.5	8.85	88.5	0.0059
14	99.984	39	86.43	64	8.08	89	0.0048
14.5	99.981	39.5	85.31	64.5	7.35	89.5	0.0039
15	99.977	40	84.13	65	6.68	90	0.0032
15.5	99.972	40.5	82.89	65.5	6.06	90.5	0.0026
16	99.966	41	81.59	66	5.48	91	0.0021
16.5	99.960	41.5	80.23	66.5	4.95	91.5	0.0017
17	99.952	42	78.81	67	4.46	92	0.0013
17.5	99.942	42.5	77.34	67.5	4.01	92.5	0.0011
18	99.931	43	75.80	68	3.59	93	0.0009
18.5	99.918	43.5	74.22	68.5	3.22	93.5	0.0007
19	99.903	44	72.57	69	2.87	94	0.0005
19.5	99.886	44.5	70.88	69.5	2.56	94.5	0.00043
20	99.865	45	69.15	70	2.28	95	0.00034
20.5	99.84	45.5	67.36	70.5	2.02	95.5	0.00027
21	99.81	46	65.54	71	1.79	96	0.00021
21.5	99.78	46.5	63.68	71.5	1.58	96.5	0.00017
22	99.74	47	61.79	72	1.39	97	0.00013
22.5	99.70	47.5	59.87	72.5	1.22	97.5	0.00010
23	99.65	48	57.93	73	1.07	98	0.00008
23.5	99.60	48.5	55.96	73.5	0.94	98.5	0.000062
24	99.53	49	53.98	74	0.82	99	0.000048
24.5	99.46	49.5	51.99	74.5	0.71	99.5	0.000037
						100	0.000029

* T-scores are S.D. values.

THE EVALUATION OF KNOWLEDGE TEST QUESTIONS

A simple method for discovering an approximate estimate of the value of multiple choice questions is described here.

First, score the papers. See the method, described in Chapter 3, of superimposing on the answer sheets an especially prepared key, with holes punched where the correct answers should appear. The total of questions answered correctly should be recorded on the answer sheet. Second, arrange the answer sheets according to scores. Third, transfer the data to tabulation sheets, prepared for this purpose. See Figure 42 for a sheet prepared to accommodate data on five questions. Much teacher time can be saved if these sheets, or similar ones of your own design, can be duplicated or mimeographed for your use. They can be used for all examinations of multiple choice type. For example, a portion of the sheet is illustrated below. The record of a student with a total score of 30 would be recorded in the step-interval of 30-31. If he selected the correct response for question 17, a tally mark would be placed in the column headed "R," or right responses. The record of a student with a total score of 29 would be entered in the step-interval 28-29. If 3 is the correct response and he selected 4, his choice would be recorded in the column headed "W,O," indicating wrong responses or omissions. Note that the correct response is shown in brackets at the bottom of the column, and the question number is placed at the top of the column. The record of a person with a total score of 22 who omitted the question is also shown in the illustration.

R	17	W,O
/	30-31	
	28-29	4
	26-27	
	24-25	
	22-23	O
	(3)	

R W,O	R W,O	R W,O	R W,O	R W,O
64–65	64–65	64–65	64–65	64–65
62–63	62–63	62–63	62–63	62–63
60–61	60–61	60–61	60–61	60–61
58–59	58–59	58–59	58–59	58–59
56–57	56–57	56–57	56–57	56–57
54–55	54–55	54–55	54–55	54–55
52–53	52–53	52–53	52–53	52–53
50–51	50–51	50–51	50–51	50–51
48–49	48–49	48–49	48–49	48–49
46–47	46–47	46–47	46–47	46–47
44–45	44–45	44–45	44–45	44–45
42–43	42–43	42–43	42–43	42–43
40–41	40–41	40–41	40–41	40–41
38–39	38–39	38–39	38–39	38–39
36–37	36–37	36–37	36–37	36–37
34–35	34–35	34–35	34–35	34–35
32–33	32–33	32–33	32–33	32–33
30–31	30–31	30–31	30–31	30–31
28–29	28–29	28–29	28–29	28–29
26–27	26–27	26–27	26–27	26–27
24–25	24–25	24–25	24–25	24–25
22–23	22–23	22–23	22–23	22–23
20–21	20–21	20–21	20–21	20–21
18–19	18–19	18–19	18–19	18–19
16–17	16–17	16–17	16–17	16–17
14–15	14–15	14–15	14–15	14–15
12–13	12–13	12–13	12–13	12–13
10–11	10–11	10–11	10–11	10–11
8– 9	8– 9	8– 9	8– 9	8– 9
6– 7	6– 7	6– 7	6– 7	6– 7
4– 5	4– 5	4– 5	4– 5	4– 5
2– 3	2– 3	2– 3	2– 3	2– 3
0– 1	0– 1	0– 1	0– 1	0– 1

Figure 42. Sample Tabulation Sheet

Since the tabulation sheet has room for five questions, it will save time if the recorder marks the answer sheets with a ruled line at the end of each five questions and records the answers to the first five questions before going on to the next five. The process of transferring the data to the tabulation sheets can be done fairly rapidly.

The number who succeeded on a question is readily obtained by counting the tally marks in the column headed "R"; and the number failing by counting each score appearing in the "W,O" column as one.

TABLE XXIII

PERCENTAGE MINIMUMS FOR FUNCTIONING OF ITEMS

Table for 3 %		*Table for 4 %*		*Table for 5 %*	
N	Minimum	N	Minimum	N	Minimum
Below 50	1	Below 38	1	Below 30	1
58-83	2	38-62	2	30-49	2
84-116	3	63-87	3	50-69	3
117-150	4	88-112	4	70-89	4
151-183	5	113-137	5	90-109	5
184-216	6	138-162	6	110-129	6
		163-187	7	130-149	7
		188-212	8	150-169	8
				170-189	9
				190-209	10

One of the first things the teacher wants to know is how difficult each question proved to be for the group. This is known as the difficulty rating of a question, and it is obtained by dividing the number of errors and omissions by the number who took the test. Thus, a question with a difficulty rating of 40 is less difficult than a question with a difficulty rating of 45. In other words, the higher the difficulty rating the more difficult the question. Teachers will have to decide for themselves what limits they will use, but they certainly would not want to retain many questions with a difficulty rating above 90 or below 10.

The teacher will want to know how the responses functioned in each question. He will not want to retain responses which were

not selected by any of the students taking the test. Sometimes it is advisable to set a limit or minimum, such as three percent, and decide to drop from further use any response not selected by at least three percent of the total number of persons taking the test. Thus, if the total number taking the test, or N, was 100, a response would have to be selected by at least three persons while if N was 75, the minimum would be two persons.

If fewer than the three responses in a question function, as arbitrarily defined by a choice of percent, then the question should be dropped or revised as it is no longer a multiple choice question.

DIFFICULTY RATING AND ITEM ANALYSIS

As an example of the difficulty rating and item analysis, take the question described earlier on trapping in soccer (Chapter 8, p. 186). When this question was given in a soccer knowledge test to forty-eight players, four chose part one. A glance at the table above will show you that this was an adequate number on any of the percentage levels listed. Eight chose part two; seventeen chose the correct response, part three. Twelve selected part four, while four selected part five. Three of the forty-eight omitted the question. Therefore, if you were using any of the percentage minimums for functioning of items listed above, you would retain all five parts. The difficulty rating, obtained by dividing the total errors and omissions ($4 + 8 + 12 + 4 + 3 = 31$) by N (48), is 65.

The worth of an item depends not only on its difficulty and the functioning of the responses, but upon its desirability for inclusion in the test as a whole, and upon its power to discriminate between students of high and low levels of general achievement in the subject involved. A question may be said to have perfect discriminating power when every student who answers the question correctly ranks higher in the scale than all students who answer it incorrectly. A question in which more students of low ability succeed than students of high ability is said to have minus discriminating power, and certainly is a poor question. Between the extremes of perfect and minus discriminating powers, questions of all degrees of discrimination are found. An inspection of the item

analysis for each question, if tabulated on a form similar to the one described here, will give you an estimate of how well the question discriminates between the various levels of ability.

INDEX OF DISCRIMINATION

According to Lindquist and Cook[6] evaluation of the degree of effectiveness of the various items must be based on a single quantitative measure which can be conveniently computed and easily compared for many items. Such a measure is referred to as the *index of discrimination,* i.e., the index of the effectiveness with which the item discriminates between individuals of different levels of information or ability. Various indices of discrimination have been studied and compared by research workers but no one index is infallible for all situations or for all groups of students, or levels of ability. Lindquist and Cook expressed a belief that the problem of securing such a measure was not solved. Likewise, Long and Sandiford,[7] while discussing the same problem, stated that the better indices differ so little in effectiveness that the selection might justifiably be made on the basis of ease of computation. An index of discrimination that has proven satisfactory for the authors is the one recommended by Swineford[10] for a heterogenous group of subjects. It is the difference between the mean total score of those persons succeeding on the item and of those persons failing on the item, or as expressed in formula form, $M_R - M_{W,O}$. It is easy to calculate, using the tabulation sheets, and since it involves determining the mean rather than the median, it is based on all the data secured. It is easier to use the median, since the median is more readily calculated than the mean, and apparently the results are quite comparable. For example, in one examination containing 86 questions an index was computed for each question using the mean, and another index using the median in the same formula. In this examination, 81 percent of the questions received the same evaluation, satisfactory or unsatisfactory, by the two forms of the index; 14 percent were rejected by use of the median and not with the mean; 5 percent were accepted by use of the median and not with the mean. Similar results have been found on other examinations. The choice then would seem

to be primarily one of preference by the teacher, with a slight advantage in favor of the median because of economy of time.

The minimum size of an acceptable index of discrimination must be determined. Any question with an index of discrimination of less than two-thirds the size of one standard deviation should be subjected to further study before it is decided to retain or drop it. The final decision on such borderline questions should be made on the basis of its importance in the content distribution and on its difficulty rating.

ESTABLISHMENT OF CRITERION SCORES THROUGH CONDUCT OF RATINGS

In conducting ratings which are to be used as a criterion for evaluating tests, it is necessary that the ability being rated be carefully defined. If it is playing ability in general, then the raters need to discuss what they consider excellent playing ability, what good, what average, and so on through the range. The size of the range depends upon the amount of discrimination that is desired. The judges should either be "experts" in their knowledge of the activity, or carefully trained students. The minimum number recommended is three. After they have met and discussed the various points on the scale, they should have time to become familiar with the chart that they are to use before the actual rating begins. The code for marking should be placed on the chart. The rating form may be subdivided into skills, such as (for volley ball): serve, set-up, pass, volley, and recovery from the net.

The players must be identified in some manner. Colored pinnies with numbers on both front and back are helpful.

The raters should work independently. The length of time that they watch the players will vary with the activity, but they should see each player active for a long enough time to be able to rank him in each of the units, and to give him a composite score.

After the judges have completed ratings and have given a score to each player, the scores should be totalled (see p. 238 for reason). Sometimes these are weighted by giving extra value to one judge's opinion, as for example the instructor of the class, who has seen

the group for a longer period of time. In this case, his ratings might be multiplied by a constant such as two.

Agreement of judges can be determined by the correlation technique. If the judges are well trained and the length of time for observation is sufficient, the coefficient should be high (at least .80).

PROCEDURE FOR CONSTRUCTING A MOTOR TEST BATTERY

The steps in construction of a motor test battery are essentially the same for all types. These steps will be outlined here and an illustration of the development of a sports battery will be given. It should be noted that it is a combination of logical planning and statistical analysis.

1. Study the Problem or Need for the Test

What tests are available which are purported to serve this particular purpose? What is the statistical evidence of their value? Are they practical to use in the space and time available? Are they designed for the age and skill of the group for which they are to be used? Are the standards suitable for this group? If the answers to most of these questions are unsatisfactory, then work on a battery would be indicated.

2. Analyze the Ability to be Measured

List the principal skills involved; rate their importance for success in the activity. If volley ball tests were under consideration, the analysis might develop as follows:

I. Skills involved	II. Importance of skills
1. Receiving the ball	
a. from service	xx
b. from across the net	xxx
c. from teammate's set-up	xxx
2. Playing the ball	
a. service	xxx
b. pass to teammate	xxx
c. volley across the net	xxx
d. spike across the net	x
e. set-up to teammate	xx

3. Footwork
   a. avoiding foul on service — xx
   b. following into court after service — xx
   c. getting under the ball — x
   d. filling opening left by teammate — x
   e. avoiding foul at center line — x
   f. jumping — x

The importance of a skill is rated according to the frequency with which it occurs in the game, the frequency with which it presents a problem to the player, and its relative significance for successful playing. From this chart it will be noted that the most important aspects of the game for most players are to be able to put the ball in play; to receive it, i.e., keep it in play and not be responsible for the ball becoming dead; and to play the ball either to a teammate or across the net. The service is rated as slightly less important among these skills because the player must share the opportunities for service with all the teammates. On the other hand, theoretically he may be expected to play on every point served by the opponents as well as his own team; actually he does not play that often, the frequency varying with his position on the floor and the style of the game.

The footwork seems to be very closely related to receiving and playing the ball, and of little significance in itself. Its relationship to service seems most important for it is often on failure in this respect that a point is lost. Opportunities for other items listed in footwork occur infrequently or are really a part of some other skill; therefore, they are ranked low. The other items of low rating occur only occasionally, or for a few players on the team rather than for the team as a whole.

3. Select the Experimental Items

The previous study of available tests may have yielded some of value, or have suggested forms which may be adapted easily. Many times there may be little available in a desirable form for the important aspects selected in step 2. In this case one's ingenuity constitutes the chief source of ideas. It is usually advisable to try out these ideas on small groups to be sure that they will work, that the dimensions and markings are satisfactory, and the scoring scheme feasible. This should lead one to a written description of

the test, the instructions to be given the students, number of trials to be used, and plans for administering and recording results.

In the example of volley ball the test to be selected or devised would deal principally with service, receiving and volleying the ball. More than one service test might be tried in an experimental stage. They might differ on the basis of different theories of a good service. In one case the players may be coached to serve into the corners or at least around the edges of the court with little consideration for the flight of the ball. Then the test would call for a net, floor markings on one side of the court giving higher values to areas near the boundary lines than to that in the center of the court. On the other hand the players may be coached to play balls into the corners with a fast ball traveling in a flattened arc. Then the test would call for a rope somewhere above the net with reduced scoring points for each ball going above the rope.

Likewise, teams may be taught to use a variety of placement on service, or to serve into weak spots on the opponents' team. Such a situation might call for a very different target area on the floor.

In the experimental stages, much time will be saved if a single test can be given, and the results recorded in such a way as to permit scoring in different ways to represent different tests. It is essential in this case to be definite on the point of aim. For example, if the area A (Figure 43) just in front of the baseline is considered the most desirable area for service then all students aim for that. By use of a scoring chart showing the actual court, each serve may be marked for its exact point of contact. It is then possible to score the test in as many ways as seem desirable. The shaded area may be subdivided into corners and center (b) or left as an undivided area (a); the space in front may be divided into zones parallel to A (c), or subdivided into zones by lines at right angles to A (d); the space behind A (e) may be scored as zero, or the whole area scored a little less than A, or subdivided into zones the same as those in front of A (f). Each player would have a score sheet in the form of a court outline. The scorer records the number of the trial as indicated in (a). It is then possible to construct new markings on the chart and score in any way desired.

The scoring plan for the badminton serve test (p. 49) was

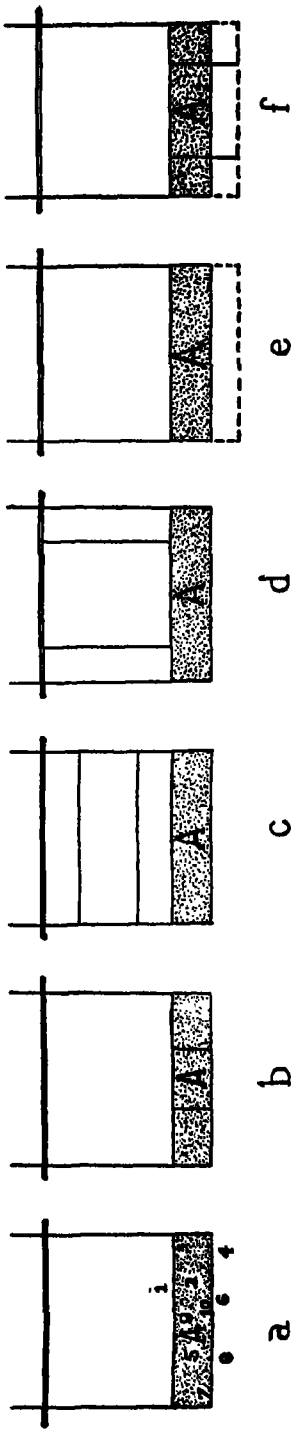


Figure 43. Suggested Target Areas for Volleyball Serve Test

devised in this way. The test was designed to test ability for a short service into the front left corner of the right service court. The scoring chart was identical with the floor markings. The student knew he was aiming for the corner of the court but knew nothing of how it was to be scored. Each service was recorded on the chart. A series of scoring schemes was tried with the eventual selection of the one given in Chapter 4 since it correlated the most highly with the criterion.

The volley ball tests to be used in this illustrative discussion will be called:

1. Service test #1
2. Service test #2
3. Volleying test #1
4. Volleying test #2
5. Set-ups
6. Net recovery

The details of each test would be determined at this stage. They are omitted here for the sake of brevity.

4. Select and Obtain the Criterion

Every test must be compared with some criterion as a yardstick of its efficacy. In rare instances the study of available tests may have shown some long or complicated measure which is otherwise acceptable. The purpose now may be to devise a short form which will give similar results with greater economy of time and effort. Under these circumstances the long form may be used as the criterion and administered to each subject. The main value of this is that the criterion is more or less objective and of known worth.

However, it more frequently happens that available measures are inadequate for such use. The best alternative, then, is to secure subjective ratings which are as good as possible. This is usually done by having at least three judges or experts rate each subject. (See p. 232 for rating procedures.) It is essential that these judges see the subjects in action in the activity, not in tests, though the ratings should be done at a time not too remote from that in which the tests are given. The scale to be used usually contains five categories,

each carefully defined. The judges should be instructed to use all five categories and not just the middle three, as frequently happens. However, knowing the nature of the normal curve it is to be expected that there will be comparatively few cases in the first and fifth categories and the greatest number in the middle one.

The various judges each give an independent rating to each subject. In utilizing these ratings it is better to use the sum of the judges' ratings than the average as the criterion. This gives equal weighting to every judge's opinion, but this would also be true if the average were used. The sum also reduces the mathematical steps by one, and thereby saves times and eliminates an opportunity for error. None of these, however, constitutes the really important reason for using the sum rather than the mean or median. If the median is used, all subjects will fall within the five point range; the same is true with the mean unless fractional values are used. When the sum is used, the range is greater. On a five point scale with three judges, the range is 3 to 15, with four judges, 4 to 20. This gives a range more nearly comparable to the test score range for computation of the correlation.

It must be kept clearly in mind that the work connected with establishment of a criterion measure is necessary only in a project to evaluate or develop a new or comparatively unknown battery. It is not part of a regular testing program.

The agreement obtained between the judges' ratings will depend upon the comparative skill and knowledge of the judges, upon the clearness of the instructions on aspects to be rated, and upon the length of time available for the judges to observe the players. Most studies yield a coefficient between .70 and .85 on agreement of judges.

In the volley ball example, the criterion could be of either form suggested here. There are volley ball batteries available which could be given as the basis for comparing new tests. Or, the judges' ratings could be made, while the various teams played. In this example, let us use a rating by three carefully selected judges. Each is provided with a detailed description of a five point rating scale. Each sees the players on three successive days of playing.

The agreement between judges as determined by a correlation of their ratings could be as follows:

Judges 1 and 2.... .79
Judges 1 and 3.... .72
Judges 2 and 3.... .80

These would be considered satisfactory. The sum of the three judges' ratings is now used for the player's score on playing ability, which is the criterion for evaluating the experimental items.

5. Select the Subjects to Be Used

The subjects used in the development of the battery should be representative of the group on which you eventually wish to use the tests. That means that validation on a college group would indicate satisfactory use on a similar college group, but not on a high school group unless a similar project also proved its worth with the younger age. The experience of the group must be considered, also. For example, different results may be obtained on beginning, and advanced groups of like age. The results may be similar enough that the battery may be used for both levels but it is best to keep the two groups separate for statistical analysis.

There is no magical number which can be given as the one which will give satisfactory results. However, as a generalization it can be said that approximately 100 should be considered a minimum. The number increases with the variability of the skill demonstrated, the lowered reliability of the measures, and the confidence which you wish to place in the results.

In the volley ball example, let us assume that the test is to be developed for beginning players in college. The subjects available include 130 cases from three classes. All will be used except the few who may be absent during the testing and rating periods.

6. Determine the Reliability of the Experimental Test Items

The ideal method of ascertaining the reliability of the separate items is to administer each in identical form on two successive days, and then to correlate the results. If the test is measuring consistently it should yield a very high coefficient (at least in the 90's); since this does not permit opportunity for practice between the two administrations, and it does not permit much change in

TABLE XXIV

RELIABILITY OF TESTS USED IN THE ILLUSTRATIVE PROBLEM OF TEST CONSTRUCTION

	# of cases	# of trials	r for repeated test	Odd-even r	Spearman-Brown r
1. Service test #1	130	10		.81	.90
2. Service test #2	130	20		.74	.85
3. Volleying test #1	126	10	.84		
4. Volleying test #2	125	10		.50	.67
5. Set-up test	120	20	.80		
6. Net recovery test	124	10	.75		

Assuming that other groups and situations are similar and that the tests are given in identical form comparable results may be expected.

TABLE XXV

COEFFICIENTS STEPPED UP BY THE SPEARMAN-BROWN PROPHECY FORMULA AS AN ESTIMATE FOR TWICE THE NUMBER OF TRIALS

Odd-even	Spearman-Brown	Odd-even	Spearman-Brown	Odd-even	Spearman-Brown
.60*	.75	.70	.82	.80	.89
.61	.76	.71	.83	.81	.90
.62	.77	.72	.84	.82	.90
.63	.78	.73	.84	.83	.91
.64	.78	.74	.85	.84	.91
.65	.79	.75	.86	.85	.92
.66	.80	.76	.86	.86	.93
.67	.80	.77	.87	.87	.93
.68	.81	.78	.88	.88	.94
.69	.82	.79	.88	.89*	.94

* For any values not included in this table compute by use of the formula on p. 241.

physical status or attitude. In general it is better than two repetitions on the same day because some will be affected more than others by fatigue, and it also gives opportunity for transitory factors of interest, feelings of physical well-being, "jinx" or "off day" performances to enter in. In general these latter factors would tend to reduce the coefficient slightly, but these factors do influence results when used on student groups; therefore, they should be given opportunity to operate when evaluating the tests.

It is not always possible to give two complete repetitions of the test, especially when it is long. In that case a randomly selected sample of about 100 subjects may be given the test twice and reliability estimated from this group. In case of the longer tests, however, it is probably better to simply correlate half the trials against the other half for each subject. This is known as the *odd-even* method. The sum of the scores on the odd numbered trials is correlated against the sum of the even numbered trials. (See Chap. 2, p. 20). This type of splitting is preferable to first and last halves as it tends to even out practice or fatigue effects.

When the odd-even method is used it is permissible to use the Spearman-Brown Prophecy formula * to step up the coefficient to twice the length in the correlation since this is the actual length of the test to be used in the remaining statistical analysis and in later administration. (See Table XXV.) Also, if the reliability is found to be too low (below .80) the Spearman-Brown formula may be used to estimate the reliability for an increased number of trials. When the number of trials is determined for the desired reliability, the lengthened form may be given to all subjects. Or, if a longer form is impractical, the test is now discarded because of its low reliability.

It will be seen then that the reliability of the test is partially a function of the length of the test. For that reason it is well to plan on several trials of any test where there is considerable variation in performance and any element of chance involved. This doubt-

* The Spearman-Brown formula is $r_x = \frac{Nr}{1 + (N-1)r}$ where

r_r = coefficient to be estimated
N = 2 or proportion of increase in length
r = correlation obtained on the halves

less explains the reason why beginners frequently require more trials than advanced players on the same test.

It must be kept in mind that there is no such thing as *the* reliability of a test. It is always for the group and under the circumstances stated. In the volley ball illustration the reliability of these tests for the 130 beginning college students is presented in Table XXIV.

7. Determine the Validity of the Test Items

The reliability should always be computed first and all experimental items with very low reliability can be dropped before attempting correlation with the criterion. This saves time, as those with low reliability invariably yield low validity coefficients and would be dropped for that reason.

The validity of the test is determined by correlating the scores of each subject with his criterion score. If previous work has been well done and the experimental items are of the right type these coefficients will range from about .60 to .80 or .85. Occasionally, a single item will be found with a coefficient in the .80's and may be considered satisfactory alone. In general, tests of this level are not high enough for individual prediction or evaluation. For that reason it is usually necessary to combine two or three tests for a battery. Those with low validity coefficients are not considered further.

In the volley ball illustration, test 4 has a reliability of only .50. It would not be practical or possible to increase its length to secure a sufficiently high reliability. Therefore, it will be dropped.

The validity of a test is also in terms of the group and circumstances under which it is established. Therefore, we will state the validity of these tests for the beginning college group of 130 to be as follows:

1.	Service test #1	.780
2.	Service test #2	.760
3.	Volleying test #1	.821
4.	Set-up test	.745
5.	Net recovery	.600

The test on net recovery will be dropped because of low validity. The best single test is volleying, and if only one test can be used that is the best selection.

8. Compute the Intercorrelation of the Experimental Items

During these successive steps the number of experimental items is gradually reduced. The low reliability and low validity items have already been discarded. Likewise, it is not helpful to try to add measures which are highly related or measure the same thing. This necessitates correlating each item retained at this stage with every other item.

In this example, the intercorrelations for the volley ball tests follow:

	1	*2*	*3*	*4*
1. Service test #1				
2. Service test #2	.70			
3. Volleying test	.25	.40		
4. Set-up test	.50	.45	.30	

As would be expected the two service tests correlate rather highly. The other tests show low but varying degrees of relationship.

9. Combine Items and Obtain Multiple Correlation with the Criterion

The logic for combining items should now be clear. Each item to go into a battery should have a relatively high validity coefficient, but it should have a minimum relationship to the other items in the battery.

The multiple correlation may be computed by the usual method referred to in any text on statistics which covers advanced correlation technique. There is a short cut to that computation which consolidates some of the intermediate steps. A sample of the computational sheet to be used is shown in Table XXVI.

The multiple correlation is a process of obtaining the best possible combination with the criterion, and the degree of relationship between that combination and the criterion. Let us use the following symbols to simplify the discussion.

R—multiple correlation coefficient
0—criterion
1—service test #1
2—service test #2
3—volleying test
4—set-up test

You are now ready to decide upon combinations which might be plausible. The two service tests are too highly related to select both, though either is good alone. Service test #1 and volleying have a low inter-correlation (.25) and, therefore, should be tried. By the same reasoning combinations with known decreases in R are as follows: 1-3, 2-3, 3-4, 2-4. Other combinations would yield an R too low to be of value.

The computation of $R_{0.13}$ is given in Table XXVI as an illustration of the short method of computation.[9] The coefficients are as follows and confirm the prediction above:

$$R_{0.13} = .970$$
$$R_{0.23} = .945$$
$$R_{0.34} = .896$$
$$R_{0.24} = .813$$

These coefficients indicate that service test #1 and volleying constitute the best combination, although the battery with service test #2 is almost as good. However, it will be remembered that service test #2 has twice as many trials as #1 and would, therefore, be impractical to use when #1 is available and of known value. The combination of volleying and set-ups also yields a very high coefficient. If the set-ups test can be given in less space and time, it might be used instead of service test #1 and without question would be chosen instead of service test #2. Hence the basis for selection is not solely the size of the coefficient; good judgment and practical assessment must be applied.

10. Compute the Regression Equations

Having decided upon the battery to be used, the method of combining the tests must be prepared. The regression equation takes into account the variability of the raw scores on each test

TABLE XXVI

SAMPLE DOOLITTLE SHEET WITH COMPUTATION OF MULTIPLE CORRELATION FOR VOLLEY BALL BATTERY*

0 = criterion
1 = serve test
2 = volleying test

	Directions	*a*	*b*	*c*	*d*	*e*	*f*	*g*
1	Insert values for r's	1.000	.250	r_{13}	r_{14}	r_{15}	r_{16}	—.780
2	Divide line 1 by —1	.000	—.250					—r_{01}
								+.780
3	Insert values for r's		1.000	r_{23}	r_{24}	r_{25}	r_{26}	—.821
4	Multiply items in Line 1, b to x, by b_{x}		—.063					—r_{02}
								+.195
5	Add algebraically Lines 3 and 4		+.937					—.626
6	Divide line 5 by negative b_{x}							+.657
7	Insert values for r's			1.000	r_{34}	r_{35}	r_{36}	—r_{03}
8	Multiply items in Line 1, c to x, by c_{x}							
9	Multiply items in Line 5, c to x, by c_{x}							
10	Add algebraically Lines 7, 8, 9							
11	Divide Line 10 by negative c 10							
12	Insert values for r's				1.000	r_{45}	r_{46}	—r_{04}
13	Multiply items in Line 1, d to x, by d_{x}							
14	Multiply items in Line 5, d to x, by d_{x}							
15	Multiply items in Line 10, d to x, by d_{11}							
16	Add algebraically Lines 12, 13, 14 and 15							
17	Divide Line 16 by negative d_{1x}							
18	Insert values for r's					1.000	r_{56}	—r_{05}
19	Multiply items in Line 1, e to x, by e_{x}							
20	Multiply items in Line 5, e to x, by e_{x}							
21	Multiply items in Line 10, e to x, by e_{11}							
22	Multiply items in Line 16, e to x, by e_{17}							
23	Add algebraically Lines 18, 19, 20, 21, 22							
24	Divide Line 23 by negative e_{xx}							
25	Insert values for r's						1.000	—r_{06}
26	Multiply items in Line 1, f to x, by f_{x}							
27	Multiply items in Line 5, f to x, by f_{x}							
28	Multiply items in Line 10, f to x, by f_{11}							
29	Multiply items in Line 16, f to x, by f_{17}							
30	Multiply items in Line 23, f to x, by f_{1x}							
31	Add algebraically Lines 25, 26, 27, 28, 29, 30							
32	Divide Line 31 by negative f_{xx}							

(This worksheet is set up for combining a maximum of six items with the criterion. It may be used for any number less than six but becomes really economical with four or more items.)

Substitute values from above table for symbols in following equations (B's for each equation found when each variable in turn is solved for) and solve equations for the regression coefficients, B_1, B_2, B_3, B_n.

$(B_4) = x_{25}$
$(B_4) = (B_4)f24 + x24$
$(B_4) = (B_4)f17 + (B_4)e_{17} + x_{17}$
$(B_4) = (B_4)f_{11} + (B_3)e_{11} + (B_4)d_{11} + x_{11}$
$(B_3) = (B_4)f_3 + (B_3)e_3 + (B_4)d_3 + (B_3)c_3 + x_3 = +.657$
$(B_1) = (B_4)f_2 + (B_3)e_2 + (B_4)d_2 + (B_3)c_2 + (B_2)b_2 + x_2 = (.657 \times -.25) + .78 = +.516$

FORMULA FOR MULTIPLE CORRELATION

Having found the several regression coefficients, the multiple R is to be found by the following formula:

$$R_{0.12\ldots n} = \sqrt{B_1r_{01} + B_2r_{02} + B_3r_{03} + \ldots\ldots + B_nr_{0n}}$$

$B_1r_{01} = .516 \times .780 = .402$

$B_2r_{02} = .657 \times .821 = .539$

$.941$

$R = \sqrt{.941} = .97$

*Worksheet quoted from Journal of Educational Research, by permission of the publishers 9.

and the relative value of each test in the total battery. The formula for this computation reads:

$$B_1\left(\frac{\sigma_0}{\sigma_1}\right)\text{test 1} + B_2\left(\frac{\sigma_0}{\sigma_2}\right)\text{test 2} + \ldots\ldots B_n\left(\frac{\sigma_0}{\sigma_n}\right)\text{test n}$$

As applied to $R_{0.18}$ in the volley ball illustration it reads

$$.516\left(\tfrac{2.5}{7.0}\right)\text{service test \#1} + .657\left(\tfrac{2.5}{10.0}\right)\text{volleying test} = .184\text{ service test \#1} + .164\text{ volleying test}$$

Likewise, the regression equations for the other batteries will read

.076 service test #2 + .154 volleying test
.177 volleying test + .189 set-ups
.085 service test #2 + .162 set-ups

However, weightings such as these are too time consuming, and fortunately unnecessary. All that is required is to get the proper proportion of each test. In these examples simple addition of the raw scores would not be far out of line but in some batteries simple addition would be completely impossible. For example, if combining the scores on a running and throwing test, most of the running scores might vary between 4½ and 8 seconds, most of the throwing scores might vary between 25 and 75 feet. Simple addition would give undue importance to the throwing score; inadequacy or superiority in running would show up only if the players received comparable scores on both tests.

In order to maintain proper proportions between the respective tests it is necessary only to compute relationship between the weightings and substitute in the equation. For example,

.184 service test #1 + .164 volleying test

would be changed to

1.1 service test #1 + 1. volleying test.

This computation is then simple enough to be done without paper and pencil or tables. It has an advantage over simple addition of keeping the scores on the two tests in proper proportion.

The simplified versions of the second battery might read

1. service test #2 + 2. volleying test

or

.5 service test #2 + 1. volleying test

The ease of computation would be about equal in those two forms.

In the third case,

1. volleying test + 1.1 set-up test

is much simpler than

.9 volleying + 1. set-up test

In the last case again it makes little difference whether you use

1. service test #2 + 2. set-up test

or

.5 service test #2 + 1. set-up test

If T-scales are computed for each test separately and the T-scores added, this sum may be used instead of a score computed by the regression equation. This is an easier process and the results are comparable.

11. Compute Norms for Use of the Tests

Raw scores are not very meaningful to either the teacher or the student. Therefore, it is considered an essential part of any project in which tests are being developed that standards of some sort be devised. A very convenient form is the T-scale. This procedure has been discussed above.

BIBLIOGRAPHY

1. Bovard, John F., and Cozens, Frederick W.: *Tests and Measurements in Physical Education,* Second Edition. W. B. Saunders Company, New York, 1938, Chapters 13, 14, 18
2. Garrett, Henry E.: *Statistics in Psychology and Education.* Longmans, Green & Company, New York, 1935, Chapter 6
3. Glassow, Ruth B., and Broer, Marion R.: *Measuring Achievement in Physical Education.* W. B. Saunders Company, 1938, Chapters 1 and 2
4. Holzinger, Karl J.: *Statistical Methods for Students in Education.* Ginn & Company, New York, 1928
5. Lindquist, E. F.: *A First Course in Statistics.* Houghton Mifflin Company, Cambridge, 1938
6. Lindquist, E. F., and Cook, Walter W.: "Experimental Procedures in Test Evaluation," *Journal of Experimental Education,* 1, 1933, p. 163
7. Long, John A., and Sandiford, Peter.: "Validity of Test Items," *Bulletin of Department of Educational Research,* University of Toronto, 3, 1935, p. 118
8. McCloy, C. H.: *Tests and Measurement in Health and Physical Education,* 2d edition. F. S. Crofts & Sompany, New York, 1942, Chapter 25
9. Peters, Charles C., and Wykes, Elizabeth Crossley: "Simplified Methods for Computing Regression Coefficients and Partial and Multiple Correlations," *Journal of Educational Research,* 3, May 1931, p. 383
10. Swineford, Frances: "Validity of Test Items," *Journal of Educational Psychology,* 27, 1936, p. 68